DOCUMENT RESUME

ED 444 730 PS 028 816

AUTHOR	Oden, Sherri; Schweinhart, Lawrence J.; Weikart, David P.
TITLE	Into Adulthood: A Study of the Effects of Head Start.
INSTITUTION	High/Scope Educational Research Foundation, Ypsilanti, MI.
ISBN	ISBN-1-57379-089-3
PUB DATE	2000-00-00
NOTE	243p.; "With Sue M. Marcus and Yu Xie. Commentaries by Suzanne M. Randolph, Mary Lou de Leon Siantz, Sheldon H. White, Edward Zigler, and Sally J. Styfco."
AVAILABLE FROM	High/Scope Press, 600 North River Street, Ypsilanti, MI 48198-2898; Tel: 800-407-7377 (Toll Free); Fax: 800-442-4329 (Toll Free); e-mail: press@highscope.org; Web site: http://www.highscope.org.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Comparative Analysis; Criminals; Early Experience; Early Parenthood; Economic Status; Educational Attainment; Employment; Followup Studies; Grade Point Average; Longitudinal Studies; *Outcomes of Education; *Preschool Education; *Preschool Evaluation; Program Effectiveness; Special Education; Welfare Services; *Young Adults
IDENTIFIERS	High Scope Preschool Curriculum Study; *Project Head Start

ABSTRACT

This report describes the Long-Term Benefits of Head Start (LTBHS) study designed to address questions regarding the long-term effects of the Head Start program on the children and families served. Following the introduction, Chapter 2 presents a review of the research on Head Start, focusing on Head Start Planned Variation research, major integrative reviews demonstrating improved school readiness of Head Start children and the elementary school fade-out effect, and reviews of non-Head Start demonstration models. Chapter 3 describes the LTBHS study's methodology and design, including the makeup of the diverse sample of 622 young adults who participated in Head Start in 1970-71; presents the strategies used to identify and locate study subjects for follow-up; and concludes with an overview of the participant interviews and data collected from school, public assistance, and criminal records. Chapter 4 explains how High/Scope researchers arrived at the LTBHS study's core variables. Chapter 5 summarizes findings on individual characteristics and family background. Chapter 6 summarizes how Head Start children, compared with non-Head Start children, fared into their young adult years. Chapter 7 summarizes the LTBHS study findings and discusses implications for future Head Start development and research. Findings indicated that some groups of study participants were better off as a result of their Head Start participation. Findings also indicated potential benefits for the Model (High/Scope) Head Start group versus the regular Head Start group. Although most findings were not of the expected magnitude or statistical significance, the pattern of findings across most of the educational and life-situational outcome variables had the same direction as the pattern of significant findings obtained in the High/Scope Perry Preschool study and the High/Scope Preschool Curriculum Comparison study. The report's six appendices include resources in locating study participants, definitions for the study's core variables, additional

Reproductions supplied by EDRS are the best that can be made
from the original document.

data tables, and additional findings. Four commentaries on Head Start and the LTBHS study conclude the report. (Contains 121 references.) (KB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 444 730

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Into Adulthood

A Study of the Effects of Head Start

Sherri Oden

Lawrence J. Schweinhart

David P. Weikart

with

Sue M. Marcus

Yu Xie

commentaries by

Suzanne M. Randolph, University of Maryland – College Park

Mary Lou de Leon Siantz, University of Washington

Sheldon H. White, Harvard University

Edward Zigler and Sally J. Styfco, Yale University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Kathleen M. Woodard
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
2B

PS 028816

BEST COPY AVAILABLE

Into Adulthood

Related High/Scope Press Publications

Significant Benefits: The High/Scope Perry Preschool Study Through Age 27

Lives in the Balance: Age-27 Benefit-Cost Analysis of the High/Scope Perry Preschool Program

Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19

Lasting Differences: The High/Scope Preschool Curriculum Comparison Study Through Age 23

Training for Quality: Improving Early Childhood Programs Through Systematic Inservice Training

Educating Young Children: Active Learning Practices for Preschool and Child Care Programs
A comprehensive up-to-date description of the High/Scope preschool model

Models of Early Childhood Education

Available from
High/Scope® Press
A division of the
High/Scope Educational Research Foundation
600 North River Street, Ypsilanti, MI 48198-2898
ORDERS: phone (800)40-PRESS, fax (800)442-4FAX
e-mail: *press@highscope.org* Web site: *www.highscope.org*

Into Adulthood

A Study of the Effects of Head Start

Sherri Oden

Lawrence J. Schweinhart

David P. Weikart

with

Sue M. Marcus
Yu Xie

commentaries by
Suzanne M. Randolph, University of Maryland—College Park
Mary Lou de Leon Siantz, University of Washington
Sheldon H. White, Harvard University
Edward Zigler and Sally J. Styfco, Yale University

HIGH/SCOPE
PRESS

Ypsilanti, Michigan

Published by
HIGH/SCOPE® PRESS

A division of the
High/Scope Educational Research Foundation
600 North River Street
Ypsilanti, Michigan 48198-2898
(734) 485-2000, FAX (734) 485-0704
press@highscope.org www.highscope.org

Copyright © 2000 by High/Scope Educational Research Foundation. All rights reserved. Except as permitted under the Copyright Act of 1976, no part of this book may be reproduced or distributed in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage-and-retrieval system, without prior written permission from the publisher. The name High/Scope and its corporate logos are registered trademarks and service marks of the High/Scope Educational Research Foundation.

Marge Senninger, *High/Scope Press editor*
Margaret FitzGerald, Metaphor Marketing, Inc., *Cover design, text design, and production*

Library of Congress Cataloging-in-Publication Data
Oden, Sherri
Into adulthood : a study of the effects of Head Start / Sherri Oden, Lawrence J. Schweinhart, David P. Weikart with Sue M. Marcus, Yu Xie ; commentaries by Suzanne M. Randolph ... [et al.].
p. cm.
Includes bibliographical references.
ISBN 1-57379-089-3
1. Socially handicapped children--Education (Preschool)--Florida--Longitudinal studies. 2. Socially handicapped children--Education (Preschool)--Colorado--Longitudiani studies. 3. Head Start Program (U.S.)--Longitudinal studies. I. Schweinhart, L.J. (Lawrence J.), 1947- II. Weikart, David P. III. Title.

LC4092.F55 O34 2000
372.21--dc21 99-054686

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

Contents

Tables xi
Preface xiii

1 INTRODUCTION 3
- **Head Start's Challenge—Then and Now 3**
- **Perspectives on Head Start Research 4**
- **About This Report 4**
  - A Review of the Head Start Knowledge Base 5
  - The Study Methodology 5
  - The Variable Selection and Data Analysis 6
  - The Implications for Future Development and Research 7

2 HEAD START RESEARCH—AN HISTORICAL VIEW 9
- **The Aims of Head Start—Who Is Served 9**
- **Specific Major Head Start Research Efforts 10**
  - Head Start Planned Variation Research 11
    - *Findings about intelligence test scores 12*
    - *Findings about model implementation 12*
  - Research Following Children Into School: The Follow Through Project 13
    - *Early findings and methodological limitations 13*
    - *Studying long-term effects 14*
  - Summarizing the HSPV Study and the Follow Through Studies 14
- **The Major Syntheses of Research on Head Start and Other Early Intervention Programs 15**
  - Two Examinations of Research on Head Start 15
    - *The Head Start Synthesis study 16*
    - *The 1997 U.S. General Accounting Office report 17*
  - Three Examinations of Research on Demonstration Programs 19
    - *The Consortium for Longitudinal Studies 20*
    - *The Center for the Future of Children report 20*
    - *The literature review of the High/Scope Perry Preschool Project 21*
  - Summarizing the Major Syntheses 22
    - *The Head Start Synthesis study 22*
    - *The 1997 GAO report 22*
    - *The three examinations of research on demonstration programs 22*
- **Some Specific Long-Term Studies of Demonstration Early Childhood Programs 23**
  - The Abecedarian Project 23
  - The Evaluation of the Chicago Child-Parent Center Programs 23
  - The Syracuse University Family Development Research Program Longitudinal Follow-up Study 24
  - The High/Scope Longitudinal Studies 24

The High/Scope Perry Preschool study 24
The High/Scope Preschool Curriculum Comparison study 25
Summarizing the Specific Long-Term Studies of Demonstration Early Childhood Programs 25
Building on the Knowledge Base—The Long-Term Benefits of Head Start Study 25

3 THE DESIGN AND METHODOLOGY OF THE LONG-TERM BENEFITS OF HEAD START STUDY 27
The LTBHS Study Groups 27
The LTBHS Study Sites and Groups 28
The Florida Site 28
The Colorado Site 29
The Study Design 30
Potential Sources of Study Bias 31
Potential Bias in the Head Start Recruitment Process 31
Potential Bias in the Data Collection Procedures 32
The Study Procedures 33
Selecting Potential Members of the Study's No–Head Start Group 33
Locating, Recruiting, and Interviewing Potential Study Participants 34
Selecting Local Project Coordinators, Locators, and Interviewers 34
Local project coordinators 34
Locators and interviewers 35
Locating and Contacting Potential Study Participants 35
The approach used to locate participants 35
The location obstacles encountered and overcome 36
The procedures used to contact participants 36
Interviewing Potential Study Participants 37
Study participants who were difficult to schedule 37
Study participants living outside the study site 38
The last-interviewed study participants 38
Interviewee compensation 39
Response Rates of the Potential Study Participants 39
The Final Study Sample 39
Collection of Study Data 41
The Interview and Adult APL Survey 41
The interview instrument 41
The Adult APL Survey 42
Interviewer training 43
Interview and Adult APL Survey data collection 43
The Records Search 44
Obtaining access to records data 44
Educational records 44
Public assistance records 45
Criminal records 45

      U.S. Bureau of the Census records 46
    Data Coding and Entry 46
      Occupational data sources 46
      Occupational codes 47
  The Data Analysis Approach 47
    The Advisory Panel and the Independent Analysts 47
    The Analytic Approaches 48
      Variables and predictions from the High/Scope Perry Preschool study 48
      A common core of key variables 48
      Statistical adjustments for original group differences 48
  Summary of the Study Methodology 49

4 THE CORE VARIABLES OF THE LTBHS STUDY 51
  Basis for Selecting the Core Variables 51
  Deriving the Core Variables From the Data Sources 52
  The Core Covariates 52
    Study Participant's Group Status 53
      Head Start group versus No–Head Start group 53
      Model Head Start group versus Regular Head Start group 53
    Study Participant's Individual Characteristics 54
      Gender 54
      Racial/ethnic membership 54
      Age at interview 54
      Kindergarten participation 54
    Study Participant's Family Background 54
      Number of Siblings 55
      Household size/structure 55
      Mother's and father's education 55
      Mother's and father's employment 55
      Family welfare assistance 56
      Site tract-level characteristics 56
  The Core Outcome Variables 56
    Study Participant's Education 57
      Educational attainment 57
      Grade point average (GPA) 57
      Years in educable mental impairment (EMI) classes 58
      Adult APL Survey Score 59
    Life Situation/Economic Status 60
      Employment/school-enrollment status 60
      Teen-parent status 62
      Public assistance 62
    Criminal Activity—Arrests and Convictions 64
  Summary 65

5 FINDINGS ON THE ORIGINAL COMPARABILITY OF THE LTBHS STUDY GROUPS 67

- **An Overview of the Analysis Process 68**
  - Selecting the LTBHS Study's Variables and Predicting Findings 68
  - Examining the LTBHS Study Groups' Initial Comparability 69
  - Adjusting for Initial Group Differences 69
- **The Findings on the Covariates: Individual Characteristics and Family Background 69**
  - Study Participants' Individual Characteristics 70
    - *Comparing the Head Start group and the No–Head Start group at the two sites 70*
    - *Summing up individual characteristics—Head Start versus No–Head Start 72*
    - *Comparing the Model Head Start group and the Regular Head Start group at the Florida site 72*
    - *Summing up individual characteristics—Model Head Start versus Regular Head Start 72*
  - Study Participants' Family Backgrounds 73
    - *Comparing the Head Start group and the No–Head Start group at the two sites 73*
    - *Summing up family background—Head Start versus No–Head Start 76*
    - *Comparing the Model Head Start group and the Regular Head Start group at the Florida site 76*
    - *Summing up family background—Model Head Start versus Regular Head Start 78*
  - The Site Tract-Level Characteristics of Study Participants' Original Households 78
    - *Comparing the Head Start groups and the No–Head Start groups at the two sites 78*
    - *Summing up site tract-level characteristics—Head Start versus No–Head Start 78*
    - *Comparing the Model Head Start group and the Regular Head Start group at the Florida site 80*
    - *Summing up site tract-level characteristics—Model Head Start versus Regular Head Start 80*
  - **Based on Background Characteristics—The Expected Pattern of Findings on Long-Term Benefits 80**
  - Comparison of the LTBHS Study and Related Studies: Participants' Backgrounds 80
    - *The Head Start Planned Variation study 81*
    - *The High/Scope Preschool Curriculum Comparison study 83*
    - *The High/Scope Perry Preschool study 83*
  - Expected Findings for the LTBHS Study 84
    - *Predictions for the Head Start versus No–Head Start groups 84*
    - *Predictions for the Model versus Regular Head Start groups 85*
- **A Summary of Expectations Based on Background Comparisons 85**

6 MAJOR FINDINGS ON HEAD START'S LONG-TERM BENEFITS 87

- **The Study Participants as Young Adults 88**
- **Estimating Head Start Benefits: The Effectiveness of Adjusting for Background Differences 88**
  - Adjusting for Measured Background Differences 89
  - Estimating Unmeasured Bias 90
  - Summary of the Adjustment/Estimation Process 92
- **The Findings on the Core Outcome Variables 92**
  - Analysis of Head Start Versus No–Head Start 93
    - *Educational outcomes 93*
    - *Economic and other life-situation outcomes 96*
  - Summary of the Head Start Versus No–Head Start Analysis 96
  - Analysis of Model Head Start Versus Regular Head Start 96

EST COPY AVAILABLE

Educational outcomes 98
Other life-situation outcomes 100
Summary of the Model Head Start Versus Regular Head Start Analysis 101
Comparing Other Specific Subgroups—Analysis by Gender 101
Educational outcomes 102
Other life-situation outcomes 102
Summary of the Analysis by Gender 104
Comparing Other Specific Subgroups—Analysis by Racial/Ethnic Membership and Socioeconomic Level 104
Educational outcomes 104
Other life-situation outcomes 105
Summary of the Analysis by Racial/Ethnic Membership and Socioeconomic Level 105
Conclusion 105

7 SUMMARY AND IMPLICATIONS 107
What Studies Have Thus Far Revealed About Head Start Effects 107
An Overview of This Study 108
The Methodology 108
The Expected Findings 109
A Summary of the LTBHS Study's Major Findings 110
Estimates of Head Start's Long-Term Effects 111
Findings for Specific Subgroups 111
Assessing the Findings 112
Implications and Recommendations 113
What Do the LTBHS Study's Findings Mean for Head Start Programs? 113
Today's Head Start: Serving a larger number and a greater diversity 113
Today's Head Start: Focusing on enhancing quality 114
Today's Head Start: Building communication/partnerships with schools and communities 115
What Do the LTBHS Study's Findings Mean for Head Start Research? 115
Continue to conduct studies of Head Start's contributions to children's development 115
Study Head Start programs that attempt to achieve enduring as well as immediate benefits 116
Conduct studies that focus on components of quality-oriented program development 116
Employ a diversity of research designs and methodological approaches 116
Summary of Lessons Learned 119
Conclusion 119

Appendix A: Directories and Other Resources Used in Locating Study Participants 121

Appendix B: The Study's Core Variables 125

Appendix C: Original Comparability of Study Groups—Additional Tables and Map 135

Appendix D: Assessment of the Long-Term Benefits of Head Start, *Yu Xie* 139

Appendix E: Endogenous Switching Regression Models, *Yu Xie* 169

Appendix F: Estimating the Long-Term Effects of Head Start, *Sue M. Marcus* 179

References 201

BEST COPY AVAILABLE

Commentaries 209
A Head Start—For Life?, *Suzanne M. Randoph* 211
The Long-Term Benefits of Head Start Study—Valuable Guidance, *Mary Lou de Leon Siantz* 217
The Futures of Head Start's Children, *Sheldon H. White* 221
The LTBHS Study's Contributions to the Future of Head Start, *Edward Zigler and Sally J. Styfco* 225

Tables

3.1 Potential and Interviewed Study Participants by Site and Group 39

3.2 Final Study Sample by Group, Site, Gender, and Race/Ethnicity 41

4.1 Core Variables Selected for Analysis 53

5.1 Head Start and No–Head Start Groups by Site, Gender, and Race/Ethnicity 70

5.2 Head Start Versus No–Head Start Groups on Study Participant Characteristics 71

5.3 Head Start Versus No–Head Start Groups on Family Background 73

5.4 U.S. Bureau of the Census Tract-Level Data on Areas Within the Two Study Sites 79

5.5 Comparison of Participants' Family Backgrounds in the LTBHS Study and Related Studies 82

6.1 Head Start Coefficients From Regression Models 95

6.2 Major Findings for Florida Study Groups 99

6.3 Florida Head Start Versus No–Head Start Females—Analysis Using Unadjusted and Adjusted Data 103

6.4 Females' Completion of High School/GED and Some Postsecondary Education in the LTBHS and High/Scope Perry Preschool Studies 103

B.1 Description and Specifications of Core Covariates 128

B.2 Description and Specifications of Core Outcome Variables 132

C.1 Study Group Background Comparability on Household Size/Structure 136

C.2 Study Group Background Comparability on Mother's and Father's Education 136

C.3 Study Group Background Comparability on Mother's and Father's Employment 137

C.4 Study Group Background Comparability on Family Welfare Assistance 137

D.1 Descriptive Statistics of Outcome Variables and Covariates 140

D.2 U.S. Bureau of the Census Tract-Level Characteristics 144

D.3 Confounding of Regional Variation With Program Variation—Grade Point Average (GPA) and Crime Outcomes in Florida 147

D.4 Logit Regression Models Predicting High School Completion 150

D.5 Logit Regression Models Predicting Attainment of Some Postsecondary Education 153

D.6 Logit Regression Models Predicting Incidence of Educable Mental Impairment (EMI) 155

D.7 OLS Regression Models Predicting GPA in Grades 1–4 157

D.8 OLS Regression Models Predicting GPA in Grades 5–8 158

D.9 OLS Regression Models Predicting GPA in Grades 9–12 159

D.10 OLS Regression Models Predicting Adult APL Survey Scores 160

D.11 Logit Regression Models Predicting Arrest Incidence 162

D.12 Logit Regression Models Predicting Conviction Incidence 163
D.13 Logit Regression Models Predicting Receipt of Public Assistance 164
D.14 Unadjusted and Adjusted Means of Outcome Variables by Head Start Status and Sex 166
E.1 Switching Regression for Years of Schooling 173
E.2 Switching Regression for GPA in Grades 1–4 174
E.3 Switching Regression for GPA in Grades 5–8 175
E.4 Switching Regression for GPA in Grades 9–12 176
E.5 Switching Regression for Adult APL Survey Score 177
F.1 Comparison of Treatment and Control Groups (Dichotomous Covariates) 181
F.2 Comparison of Treatment and Control Groups (Nondichotomous Covariates) 182
F.3 Comparison of Treatment and Control Groups (Dichotomous Outcomes) 184
F.4 Comparison of Treatment and Control Groups (Nondichotomous Outcomes) 185
F.5 Florida Head Start Versus No-Head Start—Selected Outcomes by Propensity Score Quintiles 187
F.6 Unadjusted and Adjusted Odds Ratios for Dichotomous Outcomes 189
F.7 Unadjusted and Adjusted Treatment Effects for Continuous Outcomes 190
F.8 Hidden Bias to Explain Benefits of Head Start 193
F.9 Treatment Effect Adjusted for Hidden Bias Under Selected Assumptions About an Unobserved Covariate 194
F.10 Odds Ratios by Propensity Score Quintile 196
F.11 Treatment Effects by Propensity Score Quintile 197

Preface

This book reports on the Long-Term Benefits of Head Start (LTBHS) study conducted under the auspices of the High/Scope Educational Research Foundation to examine the outcomes of the 1970–71 Head Start program in two sites—one in Florida and one in Colorado. There are good reasons why High/Scope proposed to conduct this study.

First, the two sites involved in the LTBHS study had earlier been part of the Head Start Planned Variation (HSPV) study carried out in the late 1960s and the early 1970s; High/Scope had sponsored the specific model curriculum employed in some Head Start programs at these two sites in that early, nationwide study of "model Head Start" versus "regular Head Start" participants. The HSPV findings had indicated an early positive impact for Head Start children using the High/Scope model curriculum. This new, LTBHS study, then, would in part be a follow-up on these model Head Start children in their young-adult life.

Second, High/Scope is well-known for its longitudinal studies examining the effects that preschool participation has on the lives of participants when they reach young adulthood. The High/Scope Perry Preschool Study and the High/Scope Preschool Curriculum Comparison Study are two longitudinal studies that have demonstrated that providing high-quality early childhood education programs for children living in poverty can significantly improve their educational and economic status as young adults. Thus, High/Scope's track record and expertise in conducting longitudinal research in the early childhood field helped us to garner the necessary public and private support for the LTBHS study.

The LTBHS study results that are presented in this book offer encouraging findings about the potential of Head Start to make a lasting difference and should thus be of interest to educators, program developers, and policymakers. However, this report's strongest contribution is likely to be its examination of the methodological and statistical challenges and approaches involved in the study's quasi-experimental research design (which compared a group of Head Start participants with a group who had not attended Head Start but came from the same or similar neighborhoods). *Into Adulthood* should thus be of particular interest to program evaluators and researchers carrying out either quasi-experimental or experimental longitudinal studies of Head Start and other early childhood programs.

We first began the preparations for data collection for the LTBHS study in 1986. The study was challenging from the beginning as we sought to identify for our study group those adults who had been Head Start students from the Head Start Planned Variation study of 1970–71. We also needed to form a new study group of adults who as children had lived in the same low-income neighborhoods as the HSPV participants but had not attended Head Start and had never been part of the HSPV study. To meet this challenge, we searched through teachers' old files for classroom lists and available archived computer lists of student names and addresses. Throughout the study, we received strong cooperation and practical help from the administrators and staff of Florida's Okaloosa County and Escambia County Head Start agencies and also from the administration and staff of the Head Start agency at the Colorado site. The project also benefited from the guidance of human service agencies and other leadership in the communities where the study took place. For the data collection, we were fortunate to have outstanding leadership and determination from the local study coordinators we hired for the project—Anne Armstrong, Lori Chapman, Gilda Marvray, Cenobia Moore, Ann Sitchler, Juanita Shuman, and Diane Waters.

The local study coordinators and the local staff we hired to conduct the study interviews carried out an enormous undertaking—they interviewed over 600 young-adult study participants 17 years after they had been of preschool age. The determination, dedication, and persistence of these local workers is described in this book. Staff from the schools worked diligently to collect school records data. Staff at the human service agencies cooperated with us to conduct computer searches to provide our local staff with the public assistance data. Our local staff also searched the criminal records files to collect data.

At High/Scope, members of our research staff—Jill Claxton, Kay Elder, Lyn Griffin, and Kay Long—coded and entered all the data collected. Our data analysts Ben Ben-Baruch, Ted Jurkiewicz, Zhenkui Ma, and Mei-Yu Yu organized and prepared all the data for analysis. An advisory panel of national experts—Sheldon White (chair), Mark Appelbaum, Robert Boruch, and Larry Hedges—provided outstanding analytic direction to the project's statistical analysis phase.

We were very fortunate to engage scholars Sue Marcus and Yu Xie, who provided guidance and assistance in the final data preparation and who conducted independent analyses of the data. Another scholar, Yuchin Xie, provided assistance to the analysts in the Geographic Information System coding of the study participants' original addresses.

Finally, the project was funded by the Administration for Children and Families' Head Start Bureau of the U.S. Department of Health and Human Services and by an independent foundation.

In addition to being indebted to my coauthors, to the members of the advisory panel, and to my numerous collaborators, I am grateful to the many reviewers of the early versions of this report for their very helpful suggestions. This includes Jill Claxton, Kay Elder, Ted Jurkiewicz, Amy Fowler Kinch, Judy Florian, Shannon Lockhart, Jeanne Montie, Polly Neill, Jeanne Oden, and Virginia Okoloko. I am grateful to Marge Senninger for her thorough, excellent work as the book's editor, to Pattie McDonald for her careful editorial assistance in typesetting the tables and preparing the manuscript for production, and to Margaret FitzGerald for the book's design and production.

It is my hope that this report, to which so many have contributed their efforts, demonstrates the potential contributions of research to a better understanding of the role of early education in the lives of children over time.

Sherri Oden
Principal Investigator

Technical Advisory Panel

Sheldon H. White, Panel Chair
Harvard University

Mark I. Appelbaum
University of California—San Diego

Robert F. Boruch
University of Pennsylvania

Larry V. Hedges
University of Chicago

Into Adulthood

1

INTRODUCTION

Ever since the beginning of the national Head Start program in the mid-1960s, educators, researchers, policymakers, and the general public have posed questions about the program's short- and long-term effects on the children and families it serves. This report describes a study designed to address such questions. Known as the Long-Term Benefits of Head Start (LTBHS) study, it followed up former Head Start children in their young-adult years, 17 years after their Head Start participation in 1970–71.

Both the methodology and the findings reported here should serve to advance our understanding of how to estimate the long-term effects of Head Start.

HEAD START'S CHALLENGE—THEN AND NOW

Authorized by the federal Head Start Act (most recent, U.S. Department of Health and Human Services, 1998a), the national Head Start program has been in existence for over 30 years. To this day, it has retained its original basic purposes. They are clearly described in this (April 1996) excerpt from the *U.S. Federal Register:*

> Founded in 1965, the program currently offers comprehensive services including high quality early childhood education, nutrition, health, and social services, along with a strong parent involvement focus, to low-income children nationwide. The overall goal of the program is to bring about a greater degree of social competence in preschool children from low-income families. Social competence refers to the child's everyday effectiveness in dealing with both his or her present environment and later responsibilities in school and life. It takes into account the interrelatedness of cognitive, intellectual, and social development; physical and mental health; and nutritional needs. (p. 17754)

Senator Nancy Kassebaum, in 1994, observed that Head Start's approach to fostering the relationships between children, families, and the community had become even more valid in the 1990s, in light of the rapidly changing circumstances of the American family. She pointed out that Sargent Shriver and the child development experts who formulated the Head Start approach had ". . . accurately identified the basic concepts that are needed for successful human service programs—locally controlled comprehensive services that are child-centered and family-focused" (p. 123).

The LTBHS study examines how Head Start has affected the long-term development of young people who, as children in 1970, lived in economically disadvantaged households. Any of the study's findings about the potential of Head Start, however, must be considered within a present-day context. Today, many neighborhoods and communities struggle with deepening poverty, increasing numbers of single-parent households, a diminishing number of positive role models for children, and growing despair about intervention programs having the potential to conquer poverty (Currie, 1997; Duncan, Brooks-Gunn, & Klebanov, 1994; Lewit, Terman, & Behrman, 1997; Wilson, 1987). In addition to these changes in the landscape of poverty, there are major changes taking place in the overall cultural milieu. Across the economic spectrum, increasing numbers of mothers of young children are rejoining (or never leaving) the work force, so families of all income levels seek early childhood education and care for their young children, especially for their 3- and 4-year-olds. The 1996 federal legislation requiring welfare-recipient mothers to join or rejoin the work force means that the need for early childhood education or care for the children of low-income families is especially acute (Lewit et al., 1997).

PERSPECTIVES ON HEAD START RESEARCH

Like the basic purposes of Head Start, the central question for Head Start research has remained fairly constant over the past 30 years: *Does Head Start improve economically disadvantaged children's social and cognitive development and school readiness so that they start school on an equal footing with other children?* This question is often reduced to *Does Head Start work?*

Many have argued that the central question for future Head Start research needs to be more expansive. For example, a 1990 report by the U.S. Department of Health and Human Services—*Head Start Research and Evaluation: A Blueprint for the Future*—recommended that Head Start research should focus on the diversity of Head Start programs and determine how to maximize benefits for children and families who contend with multiple kinds of circumstances. More specifically, researchers should consider that Head Start provides a diversity of local programs that are designed for economically disadvantaged children from various cultural backgrounds across all regions of the country. Expanding potential research questions further, a 1993 report by the same department—*Creating a 21st Century Head Start*—proposed that Head Start recapture its original leadership role as a national laboratory for establishing best practices in early childhood education and care.

Beyond the Blueprint: Directions for Research on Head Start's Families (Phillips & Cabrera, 1996) pointed out that the financial support for Head Start research has been episodic, resulting in periods of low activity and periods of high activity. In a preface to this same report, Sheldon White stated, "For many, understanding of the program's accomplishments had been build on intermittent discussions of the findings of a thin thread of controversial evaluation research, but studies of Head Start to date have been limited in quantity and scope" (p. *ix*). A lack of ongoing adequate support for Head Start research continues to be a major drawback to expanding the knowledge base. As recently as 1995—a time when there was an increased call from the U.S. Department of Health and Human Services for Head Start research proposals—only 0.34% of a $3.5 billion Head Start appropriation was allocated for research, demonstration, and evaluation (U.S. General Accounting Office, 1997, p. 27). Although the research to date on Head Start's effectiveness has been limited, as further noted by Sheldon White in his preface to the *Beyond the Blueprint* report, the general awareness and acceptance of the program is high.

ABOUT THIS REPORT

Whatever the direction of future research on Head Start, the importance of this report on the LTBHS study lies in what it demonstrates about designing and carrying out studies to provide true estimates of Head Start's capability to prepare young people to live and work in an ever-changing world. The remainder of this chapter provides an overview of the report.

A Review of the Head Start Knowledge Base

Chapter 2 presents a review of the research on Head Start. The review first sketches the population characteristics of Head Start, then describes the historical milestones of the research. We focus primarily on the Head Start Planned Variation research because it is the foundation for the LTBHS study and because it represents a major endeavor to conduct large-scale, national research on Head Start. As noted by Mary Kennedy (1978), there were many methodological limitations in the Head Start Planned Variation studies that evaluated the effects of continuing several Head Start curriculum models into kindergarten through third grade. However, these early studies did lay the foundation for the program research and development that extended from the 1970s to the present, including development of the Head Start Program Performance Standards for evaluating programs. These early studies also brought out the importance of using multiple measures and comparison or control groups in Head Start research.

Next, to examine the methodological issues and the major patterns of findings, Chapter 2 examines the major integrative reviews of Head Start research. These reviews demonstrate that across a variety of research approaches and outcome measures, Head Start has been found to improve children's readiness for the early school years. However, a corollary to the original research question *Does Head Start work?* often has been *Do the effects of Head Start last?* As summarized in Chapter 2, most studies of Head Start programs found what has been called a fade-out of the early effects. Specifically, over the course of the elementary school years, Head Start children's performance on various school-achievement tests has been found to eventually decline to the levels of comparable children who were not Head Start participants.

Attempts have been made to explain the fade-out. For example, Donald Campbell and Peter Frey (1970) and others (e.g., Barnett, 1992; Powell & Roane, 1996) have pointed out, in discussing compensatory education, that unless subsequent schooling continues the stimulation provided by the Head Start intervention, some fade-out of early effects should be expected. In support of this view, Lee and Loeb (1995) found that Head Start "graduates" tend to subsequently enroll in the poorest of elementary and middle schools.

Other studies have looked beyond the question of elementary school fade-out. Longitudinal studies of early childhood program models—such as those studied in the High/Scope Perry Preschool Project (Schweinhart, Barnes, & Weikart, 1993) and in the High/Scope Preschool Curriculum Comparison Study (Schweinhart & Weikart, 1997)—have followed preschool participants from low-income families into their young-adult years to find that high-quality preschool can have resilient, lasting positive effects. In the first of these two studies, the low-income children who had participated in a preschool program grew up to achieve higher levels of education and greater economic viability than did similar children who had not participated in a preschool program. Thus, along with reviewing the research on Head Start, Chapter 2 also reviews research on some non–Head Start demonstration models of early childhood programs (besides the Planned Variation models).

The Study Methodology

The methodology of the LTBHS study presents issues and provides insights for future Head Start studies. Therefore Chapter 3 describes in careful detail the LTBHS study's methodology and design, including the make-up of the large and diverse sample of 622 study participants. The sample comprises a group of Head Start children and a group of No–Head Start children in two different regions of the country; they come from a variety of urban and rural areas and include primarily three major ethnic groups—African Americans, Hispanic Americans, and Whites. In both design and methodology, the LTBHS study closely parallels the High/Scope Perry Preschool research, especially its age-19 follow-up study (Berrueta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984). However, the High/Scope Perry study's design was experimental, whereas the design of the LTBHS study is quasi-experimental.

In leading up to discussing the design of the LTBHS study, Chapter 3 first explains how the sites and the study groups of the Head Start Planned Variation study (Smith, 1973) are involved in the LTBHS study. The original Planned Variation samples, which form the major part of the Head Start groups in the LTBHS study, are described along with the LTBHS study's newly constructed No–Head Start groups. Sources of potential selection bias and other types of study bias, because they are a major challenge presented by this Head Start study, are also fully explored.

Chapter 3 next presents the strategies used to identify and locate study subjects for follow-up in this longitudinal study. With the hope that it may prove useful for similar future Head Start research, we describe specific procedures for study participant identification, location, and interviewing—both in the chapter and in Appendix A. Chapter 3 concludes with an overview of the content of the study participant interviews and of the data that researchers collected from kindergarten through postsecondary school records, public assistance records, and criminal records, as applicable.

The Variable Selection and Data Analysis

Chapter 4, along with Appendix B, explains how High/Scope researchers arrived at the LTBHS study's core variables. In a longitudinal study such as this one, it is useful to include a wide enough range of variables to provide a potential database for scholars from a diversity of specialities. Thus, the LTBHS study's database is extensive, but the study analyses focused only on those core variables that would best address the study's central research question: *Do former Head Start children equal or surpass their No–Head Start counterparts in their school years and in their young-adult years in educational achievement, economic viability, and life adjustment?* The High/Scope Perry Preschool Project's age-19 and age-27 follow-up studies (Berrueta-Clement et al., 1984; Schweinhart et al., 1993)—which addressed the same type of question and employed assessments highly similar to those used in the LTBHS study—served as the foundation for this study's variable selection and formulation.

As Chapter 4 will relate, a particular challenge in working with this study's longitudinal data set was to form composite variables from the various data sources (e.g., from educational levels reported in the interview and school records data, from self-report delinquency and criminal records data). In selecting and constructing the core variables, High/Scope researchers consulted with two independent scholars, statistical experts Yu Xie (of the University of Michigan) and Sue Marcus (of the University of Pennsylvania).

These same two scholars conducted the LTBHS study's data analyses, which they report in detail in Appendixes D, E, and F. The various data analysis methods they used—including both the traditional and the more innovative ones—were initially recommended by an independent panel of statistical and developmental experts that was chaired by Sheldon White (of Harvard University) and consisted of members Mark Appelbaum (of the University of California at San Diego), Robert Boruch (of the University of Pennsylvania), and Larry V. Hedges (of the University of Chicago). The Xie and Marcus reports, in addition to explaining their various approaches to data analysis, examine the LTBHS study's methodological issues and present and interpret the major findings.

Chapters 5 and 6 integrate and summarize the results of the analyses conducted by Xie and Marcus. Chapter 5 summarizes the findings on study participants' individual characteristics and family background, including the site characteristics connected with their childhood place of residence. Chapter 6 summarizes how Head Start children, compared with No–Head Start children, fared into their young-adult years. Also, throughout Chapters 5 and 6, the LTBHS study's findings are compared with those found in the High/Scope Perry Preschool study and other relevant studies.

The Implications for Future Development and Research

Chapter 7 summarizes the LTBHS study findings and discusses their implications for future Head Start development and research. For example, the wide-ranging young-adult circumstances that this study found occurred in both the Head Start and the No–Head Start sample alike—both groups had some members who, 17 years after preschool age, had become teachers, welfare recipients, college students, prison inmates, and so on. This finding suggests that factors in addition to Head Start participation contribute to long-term outcomes. The implication is that it is important for future research to gather data on family income levels and on family attitudes about education and childrearing approaches, as also suggested by Reynolds (1992) and others. The educational attainment and life-adjustment findings of the LTBHS study suggest many other directions that future Head Start research might take. One important direction would be to explore the relationship between developing high-quality programs and producing lasting effects on program participants.

The LTBHS study findings also suggest directions to take in developing methodology for future Head Start research. From this study's results, we can learn much about how to select appropriate samples, form study groups, select study designs appropriate to various research questions, develop valid measures of socioeconomic background, and assess site population characteristics. We can also learn something about selecting valid and appropriate outcome measures, particularly with respect to educational achievement, economic viability, and life adjustment. The data analysis work by Xie and Marcus presents instructive models for effectiveness studies and for analysis of long-term data, especially data from studies that have quasi-experimental designs. The analytic approaches of Xie and Marcus should be of interest not only to those studying the Head Start program but also to those studying interventions of other kinds—in education, health, social service.

Overall, as suggested in this book's four commentaries by Suzanne Randolph, Mary Lou de Leon Siantz, Sheldon White, and Edward Zigler and Sally Styfco, the LTBHS study moves Head Start research a step forward. Despite its quasi-experimental design and despite some other limitations inherited from the earlier Head Start Planned Variation study, the study described in the following chapters demonstrates the potential for assessing the long-term benefits of participating in a Head Start program.

2

HEAD START RESEARCH—AN HISTORICAL VIEW

> The empirical literature thus delivers good news and bad news. The bad news is that neither Head Start nor any preschool program can inoculate children against the ravages of poverty. Early intervention simply cannot overpower the effects of poor living conditions, inadequate nutrition and health care, negative role models, and substandard schools. But good programs can prepare children for school and possibly help them develop better coping and adaptation skills that will enable better life outcomes, albeit not perfect ones. The current disillusion with Head Start stems from overselling its benefits, but the evidence gives real reason not to undersell them. (Zigler & Styfco, 1994, p. 129)

As discussed by Zigler and Styfco in the *American Psychologist* article that yielded the above quotation, criticisms of Head Start are often formulated on the basis of unrealistic expectations, inappropriate interpretation of evidence, or an incomplete knowledge base. This chapter presents a review of the Head Start knowledge base, with particular attention to research that has examined the effectiveness of Head Start for the population it serves. First, the purposes and growth of Head Start are briefly described. Next, we survey the historical course of research relevant to Head Start programs. This research survey covers major individual studies as well as compilations and syntheses of a number of individual studies. It also includes research involving demonstration models of Head Start and other early childhood education programs, with particular attention to studies that are most relevant to the Long-Term Benefits of Head Start (LTBHS) study. The chapter concludes by first summarizing the methodological issues that are brought up by the reviewed research and then presenting the conceptual framework of the LTBHS study.

THE AIMS OF HEAD START—WHO IS SERVED

Head Start, which serves primarily 3- to 5-year-old children from low-income families, is not a single-focus program. It does not simply aim to improve children's school readiness or to promote parents' involvement in schools. Rather, Head Start sets both of these goals and additional ones as well. Its comprehensive goals have been summarized in many reports (e.g., Kassebaum, 1994; Zigler & Styfco, 1994; Zigler & Valentine, 1979). The overarching goal of Head Start, as described in the quote on p. 3 of Chapter 1, is "to bring about a greater degree of social competence in preschool children from low-income families." In this context, the term *social competence* means the overall developmental competence in social, cognitive, linguistic, emotional, and physical domains that is necessary for school- and life-readiness. In addition, Head Start provides many services to families—by linking Head Start children's parents to needed social services, fostering their parenting skills, providing them with literacy and job training, and encouraging their participation in program policy decisions. Head Start also provides health screening and immunizations for children, emphasizes the importance of nutrition, and provides services to children with disabilities.

According to the *Head Start 1998 Statistical Fact Sheet* (U.S. Department of Health and Human Services, 1998b), in the approximately 1,456 Head Start programs across the nation, 36% of enrolled children are African American, 31% are White, 26% are Hispanic American, 4% are Native American, and 3% are Asian American. Approximately 13% of Head Start children have disabilities. Eligibility for Head Start enrollment is based largely on the poverty threshold established each year by federal guidelines; 10% of Head Start children may come from families with an income above the poverty line. In fiscal year 1997, 61% of Head Start families had an annual income of less than $9,000, and 78% had an annual income of less than $12,000.

A recent U.S. General Accounting Office report (1997) on Head Start tracked the enrollment of children from 1965 on. From 1965 to 1970, Head Start was largely a summer program serving approximately 500,000 to 700,000 children per year. In the early 1970s, Head Start changed over to being a full academic-year program. According to the *Head Start 1998 Statistical Fact Sheet,* the program served approximately 400,000 children per year from 1971 to 1983; this figure had increased to about 540,000 by 1990, to about 620,000 by 1992, and to over 793,800 by 1997. It has been estimated that Head Start currently enrolls 30% to 35% of 3- and 4-year-old children who meet the federal guidelines for Head Start eligibility (U.S. General Accounting Office, 1997, p. 17).

Since its early beginnings, Head Start has aimed to serve the most impoverished of low-income families, and one national study of Head Start in the 1980s demonstrated that Head Start was largely succeeding in this objective (Love & Grover, 1987). The study's telephone survey of a nationally representative, stratified random sample of Head Start programs identified the most frequently employed program-enrollment criteria. The criterion that the family had low income was cited most often, followed by such other criteria as these—child had a disability; the child was of appropriate age; the family received federal Aid to Families of Dependent Children (AFDC); there were family problems; the child had been enrolled the previous year; or the family had a sibling in the program. A further in-depth assessment verified that the families enrolled in the Head Start programs, compared with the non-enrolled ("waiting list") families, had lower incomes and were more likely to be receiving some form of public assistance; their households were also more likely to be single-parent households or to have unemployed adults.

Having described the *extent* of the Head Start program to date, we next discuss the accumulated studies focused on determining how *effective* Head Start has been. The next four major sections of this chapter track the course of Head Start–relevant research, describing the findings and methodological approaches of various studies conducted since the 1960s and revealing that so far, the accumulated data on Head Start and related early childhood programs for economically disadvantaged children point to some consistent patterns of findings. The studies that reported these data, however, had to contend with a number of methodological limitations and challenges. The major focus of our review is on studies most relevant to the LTBHS study—the ones that primarily sought to examine children's educational and life outcomes after preschool participation.

SPECIFIC MAJOR HEAD START RESEARCH EFFORTS

Almost from its inception as a summer program, Head Start became a focus of evaluative research. Accounts of major Head Start studies began with the now well-known Westinghouse Learning Corporation and Ohio University report (1969).

The Westinghouse study found that the Head Start summer experience had no effect on children's intelligence test scores or affective measure scores, but later reanalyses showed promising social developmental gains (see Brown, 1978). However, despite critiques pointing out the limitations of the study's methodology and emphasizing the more positive results of the later reanalyses, the initial, "no effect" finding of this early evaluation of Head Start's summer program has haunted Head Start research ever since.

Another major investigation of Head Start was the national Head Start Planned Variation (HSPV) study of the early 1970s, mentioned in Chapter 1. This research, which is especially relevant to the LTBHS study, is examined next, followed by a brief description of research on the national Follow Through program that extended the HSPV models into early elementary school.

Head Start Planned Variation Research

The Head Start Planned Variation study, conducted from 1969 to 1972, was a part of the U.S. Office of Education's early-elementary Planned Variation Follow Through Project, which began in 1967 and continued into the 1970s. The national evaluation of HSPV involved approximately a dozen child-development-based and learning-theory-based curriculum models that were employed across 37 different Head Start sites (see Bissell, 1971; Datta, McHale, & Mitchell, 1976; Weisberg, 1973).

The HSPV curriculum models differed in focus. Some emphasized children's behavioral learning; others concentrated on their academic skills; still others were concerned with developing children's interests and initiative, active construction of knowledge, cognitive abilities, and social-emotional development. One program focused not as much on child development as on enhancing parental participation. (For a summary, see Smith, 1973; Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977.) To examine the effects of these theoretically based curriculum models, the HSPV design compared the various "model" Head Start programs with "regular" Head Start programs (i.e., programs using curricula formulated within local communities, following general Head Start guidelines). The two groups of programs were compared on a number of variables, including child outcomes and program implementation.

The stated purpose of the HSPV study was not to determine a best model, but to demonstrate a range of model approaches that could be effective in enhancing children's early learning and development. Fall (pretest) and spring (posttest) assessments were typically employed. They included determining family-background characteristics (e.g., mother's education, ethnic background), assessing child development (e.g., social, cognitive, language, and motor competencies), assessing intellectual capability and achievement (i.e., with the Stanford-Binet Intelligence Scale and the Caldwell Preschool Inventory), and assessing Head Start classroom practices (e.g., through implementation ratings and other observational assessments).

The 1970–1971 HSPV study cohort included 12 curriculum models employed in two to three study sites each (see Stebbins et al., 1977). For each curriculum model, staff training and implementation were directed by the model's designers, who were the designated study sponsors. The evaluations of program implementation and of child outcomes were independently conducted.

Because of incomplete child- or classroom-level data, only two sites per model were involved in all phases of assessment and in the final analyses. Two sites using a High/Scope model curriculum were included in the assessment of 1970–71 HSPV child outcomes. (For summaries of the High/Scope model's curriculum approach, see Hohmann, Banet, & Weikart, 1979; Hohmann & Weikart, 1995.) One of the High/Scope sites was in northwestern Florida, and the other was in northern Colorado. At these two sites, researchers found that compared with children in the regular Head Start programs, children in the

High/Scope model Head Start programs made significantly greater gains in Stanford-Binet intelligence test scores (Smith, 1973). Furthermore, the implementation assessment indicated that the High/Scope model had been effectively implemented (Lukas & Wohlleb, 1972). It was because of these positive HSPV study findings that the former High/Scope Head Start and regular Head Start children from these two study sites were followed up approximately 17 years later in the LTBHS study, to assess the course their lives had taken following the Head Start years, up until young adulthood. (Specific details on how the HSPV and LTBHS study samples and methodologies are related are presented in Chapter 3.) Among the general findings from the HSPV study, which were often inconclusive or mixed, there were some interesting effects for several specific models. Some of these general findings are described next, with special emphasis on the findings relevant to the High/Scope model.

Findings about intelligence test scores

The Stanford-Binet Intelligence Scale (Terman & Merrill, 1960) was administered in the fall and the spring for each new HSPV study cohort. The study analyses focused on the fall-to-spring gains (across the academic year) for model Head Start and regular Head Start program children. In years subsequent to the study's first academic year (1970–71), data were missing because of variations in reporting of test scores, use of alternative tests, and attrition of study subjects. All of this somewhat limited the longitudinal-assessment capacity of the HSPV research. The findings of these HSPV analyses, nevertheless, are especially relevant to the LTBHS study. At the fall assessment (pretest), across the various HSPV models and sites, the model Head Start groups' average *intelligence test scores* ranged from 77 to 100 points, and those of the regular Head Start groups ranged from 83 to 102 points. At the spring assessment (posttest), a multivariate analysis comparison of the model Head Start versus regular Head Start groups found the High/Scope model Head Start groups to have significantly higher *intelligence test scores* (Smith, 1973, p. 200 and Table VII–14). The High/Scope groups had a significant gain of 23.5 (*SD* = 12.4) points, compared with 7.2 (*SD* = 7.8) points for the regular Head Start groups ($p < .001$).

Regarding the two High/Scope study sites, the Smith (1973) report concluded:

> Thus, we estimate that the "true gain" for the southern rural site [a High/Scope site] is roughly 20 points, while the "true gain" for the northern urban site [a High/Scope site] is roughly 12 points. (p. 207)

The Smith report later provided findings that were more specific:

> One model (High/Scope) was clearly more effective than other models in producing gains on the Stanford-Binet. We estimate that "true" gains for children in the High/Scope model averaged roughly 12 to 15 points while "true" gains for the other models averaged 2 to 4 points. (p. 230)

Findings about model implementation

In addition to assessing HSPV child outcomes, researchers also assessed how well each of the HSPV models was implemented. Part I of *Implementation of Head Start Planned Variation: 1970–71* (Lukas & Wohlleb, 1972) reported that although all models were assessed as at least moderately well implemented when measured against the criteria of the various model sponsors, no model was rated as fully implemented. The independent observational assessments showed, both in the fall of the school year and again in the spring, that High/Scope's program was a model that was well implemented. In Part II of their report, Lukas and Wohlleb presented and discussed the classroom observational data, the collected records data, and the data from the interviews of site staff who participated in the study. Lukas and Wohlleb's summary of the factor analysis that was conducted by site (Appendix F, pp. 5–9) indicates that the evaluators looked at the following factors:

- *Community and teacher support* for the model (from satisfaction and efficiency ratings)
- *Sponsor support to the Head Start center* (from measures of frequency and degree of program sponsor support)
- *Staff turnover* (of teachers and aides)
- *Atmosphere of the Head Start center* (from assessments of staff friction, children's attendance and turnover rates, and the rapport between administration, staff, and children)
- *Physical facilities* (from assessments of the indoor and outdoor facilities)

The results of the implementation assessment indicated that sites using the High/Scope model had stable staffs. However, the implementation at the two High/Scope sites had some apparent limitations: While the Florida site had sufficient sponsor support, the Colorado site was not adequate on this factor—it had only 2.2 days of preservice training, compared with the Florida site's 10 days (Lukas & Wohlleb, 1972, p. 70). This is interesting in view of the fact that the Colorado High/Scope Head Start children, on the Stanford-Binet Intelligence Scale, gained fewer points than their Florida counterparts did. Lukas and Wohlleb suggested that effective implementation of the High/Scope model in Colorado had been more difficult to initiate and/or maintain, given the fact that there was less preservice training in Colorado.

According to a 1978 review of the HSPV research by Mary Kennedy (from the Office of Educational Research and Improvement, Department of Education), such implementation difficulties were typical across Planned Variation's Head Start and subsequent Follow Through (early elementary) sites. Kennedy concluded that despite having to assess child outcomes of models that were less than fully implemented, the Planned Variation research made many contributions to program development and research. The overall pattern of findings from the HSPV and Follow Through research is summarized in the next section.

Research Following Children Into School: The Follow Through Project

As explained earlier, the HSPV study was a part of the longer-term Planned Variation Follow Through Project, which was designed to provide economically disadvantaged children with a kindergarten-through-grade-3 program. The underlying assumption of the project was that a continuation of the educational approaches implemented in the Head Start model programs would be needed to preempt any potential decline of the initial Head Start effects in the elementary school years. During its 1967–95 existence, the federally funded Follow Through Project never grew beyond demonstration status—it never served more than a small fraction of the nation's former Head Start children. However, the Follow Through Project did serve as a proving ground for some 20 different early-elementary model programs, including High/Scope's early-elementary program.

Early findings and methodological limitations

The national evaluation of Follow Through also produced findings about the relative effects of model programs (see Kennedy, 1978; Seitz, Apfel, & Efron, 1978; Smith, 1973). One of the general conclusions of the research was that program outcomes varied more across sites than across curriculum models. In general, compared with their peers in regular elementary school programs, students in several Follow Through model programs made immediate, significant gains on a range of variables, such as standardized *achievement test scores* and *measures of self-esteem* and *achievement responsibility* (Stebbins et al., 1977).

As in many Head Start studies, selection bias in the Follow Through study groups cannot be ruled out, both between Follow Through and comparison groups and within and across the sites at which various curriculum models were employed. Children were not randomly assigned to study groups, and the comparison groups (in schools that agreed to participate) served as no

BEST COPY AVAILABLE

more than rough matches to the Follow Through groups.

Studying long-term effects

Despite the methodological limitations of Follow Through, several programs conducted long-term follow-up assessments of their participants. For example, a decade later, the Direct Instruction model program students in four sites were compared with elementary school students who were comparable in ethnicity and percentages of families receiving welfare assistance (Gersten & Keating, 1987). Without random assignment to groups or more-detailed information on program participants' background characteristics, it is not possible to conclude that this study's obtained student-outcome differences were due to the effects of Direct Instruction rather than to differences in group composition. Methodological limitations such as these should be kept in mind when considering that the study found former Direct Instruction students to significantly outperform comparison students at specifically selected sites in the following ways:

- At two of the sites, Direct Instruction students had a higher rate of *high school graduation.*
- At three of the sites, Direct Instruction students scored higher on ninth-grade *achievement tests.*
- At two of the sites, Direct Instruction students had less *grade-retention.*
- At one of the sites, Direct Instruction students included a lower percentage with 10 or more *school absences* per year.

There was also follow-up assessment of the High/Scope Follow Through program from the late 1970s to the early 1990s. In the early years of Follow Through, High/Scope elementary school students, compared with their peers in "as usual" elementary school programs, were not found to perform significantly better on achievement tests (Stebbins et al., 1977). By 1988–91, however, High/Scope classes in three elementary schools, versus comparison classes in the same schools and in similar schools, scored significantly higher in 22 of 40 school-achievement composite-score comparisons; in none of the 40 comparisons did they score significantly lower; over the 40 comparisons, the average High/Scope advantage was 13.2 normal-curve-equivalent points (Schweinhart & Wallgren, 1993). This evidence suggests that in the later years of the High/Scope program, High/Scope Follow Through classes improved *school achievement* more effectively than did regular elementary school classes serving otherwise comparable children. Furthermore, consistent with the High/Scope Follow Through K–3 program aims, the High/Scope students also made gains in *initiative, social relations,* and *creative skills.*

Another long-term follow-up assessment of some Follow Through participants was the Long-Term Effects of Head Start study, reported by Copple, Cline, and Smith (1987). This study included participants of model Follow Through programs in the Philadelphia School District. The investigators examined 10 years of data on participants in the Head Start/Follow Through program models in 33 schools from 1970–79. Attrition—which was up to 70% in some subgroups—was a major limiting factor. The findings indicated that having participated in Head Start/Follow Through had little or no lasting effect on achievement test scores. However, the study replicated two findings of other Head Start studies: Children who had attended Head Start (meaning Head Start/Follow Through versus Follow Through only) were less frequently *retained in a grade* during the elementary school years. Furthermore, the study found that the Follow Through children who had been Head Start children had higher *attendance rates* in their school years.

Summarizing the HSPV Study and the Follow Through Studies

The Head Start Planned Variation study and subsequent Follow Through studies found that many *model Head Start programs were effective in boosting general school readiness or intelligence test scores. But this boosting effect varied within models—by the nature of the study site and by the quality of each site's implementation of the model. Related to this, another major limitation of the research (discussed in greater detail in Chapter 3 with respect to the LTBHS study) was that the regular and model Head Start program children often were not sufficiently comparable on socioeconomic characteristics or other kinds of site characteristics. This major research undertaking, however, was instructive to later Head Start program development and research.*

THE MAJOR SYNTHESES OF RESEARCH ON HEAD START AND OTHER EARLY INTERVENTION PROGRAMS

Throughout the history of Head Start research, a number of investigators have developed integrative reviews of the body of research on Head Start, some of which have included meta-analyses of the data generated by the research. In this section, we first examine the range of studies, reviews, and other examinations of research on Head Start. Next, we examine the larger base of studies that have focused on Head Start as well as on other model programs specifically designed to provide early education and care services for children living in poverty.

Two Examinations of Research on Head Start

Several major reports have used various approaches to compile and review research on Head Start. One early compilation of Head Start research was conducted by Mann, Harrell, and Hurt (1978) of the Social Research Group at George Washington University. These investigators compiled over 700 reports on early intervention, including reports on Head Start, from computer searches of a number of databases (e.g., *Educational Resources Information Center (ERIC), American Psychological Association, Dissertation Abstracts International, Social Science Citations*, and various medical and government databases). They focused on 59 reports that investigated Head Start effects. Of these, 28 were dissertation studies, and 18 were national evaluations, largely from the Head Start Planned Variation, Home Start, and Educational Testing Service (ETS) Head Start studies; the rest of the reports included literature reviews or other documents. In their summary of Head Start's impact on children's cognitive development, social development, and health and on the family and community, Mann and his colleagues found that about half of the 28 dissertation studies on Head Start reported positive effects, and the others reported no effects; each of the national evaluations reported at least one positive program effect.

One of the major reviews in the 1980s was an integrative examination of Head Start that resulted in a 1985 report titled *The Impact of Head Start on Children, Families, and Communities: Head Start Synthesis Project* by McKey and colleagues (we will refer to this review as the Head Start Synthesis study, or simply the Synthesis study). These investigators identified the studies to be included in their synthesis from their compilation of research on Head Start (see CSR, Inc., 1995, *An Annotated Bibliography of the Head Start Research Since 1965*). Other compilations of Head Start research have been conducted more recently by Ellsworth Associates, (1995, *Head Start Research from 1985–1995: An Annotated Bibliography*) and by Abt Associates, Inc., with Yale University's Bush Center in Child Development and Social Policy (1995, *Review of Research on Supportive Interventions for Children and Families*).

A more recent (1997) U.S. General Accounting Office report on Head Start research (hereafter we refer to it as the 1997 GAO report) was based on the reports reviewed in earlier reviews and on an examination of updated computer searches of additional databases (e.g., *ERIC, Dissertation Abstracts International, U.S. Government Printing Office, Social SciSearch, Sociological Abstracts*). In contrast to other reports that largely employed a descriptive bibliographical approach, the Head Start Synthesis study (McKey et al., 1985) and the 1997 GAO report used two different integrative approaches. These two reports, because they include methodological implications for the LTBHS study, are summarized next.

The Head Start Synthesis study

From the more than 1,800 research reports reviewed in the Head Start Synthesis study (McKey et al., 1985), the investigators selected 76 studies that contained information usable for calculating effect sizes, since this would be integral to the meta-analysis they intended to employ. In the Head Start Synthesis study, the statistical effect sizes compared the performance of two groups in terms of standardized scores (the mean score for the program group [the Head Start group] minus the mean score for the no-program group was divided by the standard deviation for the no-program group). This is but one of the various meta-analytic methods that can be employed to analyze data across a number of studies and thereby gain an overall assessment of the knowledge in a particular area (Greenwald, Hedges, & Laine, 1996; Hedges & Olkin, 1985).

Study selection and the meta-analysis. The Head Start Synthesis study selected for inclusion those Head Start studies that (1) provided findings about program effects on child development or parent participation, (2) employed comparisons of Head Start versus no–Head Start groups over time, and (3) provided adequate statistical information for computing effect sizes. This included studies that varied in methodological rigor but that were judged to be otherwise appropriate for the meta-analyses. Many of the Head Start studies considered for inclusion in the Synthesis study had certain limitations that resulted in their elimination (e.g., they involved only summer programs, they involved Head Start demonstration projects rather than regular Head Start programs, they did not report primary data).

Next, the investigators categorized the selected studies with regard to their study designs, their possible sources of bias, and their methodological execution and ultimately compared their calculated effect sizes. The *pre-post* or *treatment-control* studies they reviewed assessed study participants' cognitive and intelligence test scores as well as their school readiness and achievement test scores. Across these studies, the time of follow-up assessment varied: 58 studies examined the impact of Head Start by the end of the Head Start year, 19 studies assessed impact 1 year after Head Start, 13 studies assessed impact 2 years after Head Start, and only 9 studies assessed impact after 3 or more years. (Some studies are included more than once in these category totals because they conducted assessments at several different time intervals.)

In their meta-analysis, the Head Start Synthesis study investigators reported Head Start effectiveness across a number of variables. In their analysis and presentation of the results, they separated the *treatment-control* studies from the *pre-post* studies (which lacked a control for maturation). For both types of studies, they found immediate (occurring in or by the end of the Head Start year) "educationally meaningful" effects on Head Start study participants' *global cognitive development, intelligence test scores,* and *school readiness and achievement test scores* (effect size above .25). Although the *pre-post* studies showed that the immediate effects on all these variables persisted through the end of the following year, without a comparison group, it is hard to assess how much of the effects would be due to maturation and how much to the Head Start experience. By contrast, the *treatment-control* studies provided evidence that when Head Start children are compared with no–Head Start children in subsequent school years, the positive program effects continually drop off until there are no differences, or sometimes, until there are differences favoring the comparison group children.

Only a few studies in the Head Start Synthesis study focused on variables other than those listed in the previous paragraph. In three Head Start versus no–Head Start studies, the meta-analysis found a pattern of positive and sizable Head Start effects on two other variables—Head Start participants had lower *grade-retention rates* and fewer *special education placements.* Two of these studies followed up Head Start participants from 1 to 5 years later (Bee, 1981; Cawley & Goodstein, 1966), and one study—the Rome (Georgia) Head Start study (Monroe & McDonald, 1981)—followed them up to age 18. These studies found Head Start effects on social development variables immediately after the program, and these effects persisted for 1 to 2 years before dropping off. These studies also found similar but less sizable immediate effects on Head Start children's *achievement motivation,* but these effects were no longer evident 1 to 2 years later.

Interpreting the meta-analysis findings. The potential of the Head Start Synthesis study was limited by the quality of the 76 studies selected for synthesis. The major methodological limitations (especially for conducting meta-analyses) identified by the Head Start Synthesis study investigators are the following:

- Pre-post designs without comparison groups to control for maturation
- Attrition in the longer-term studies
- Small sample sizes
- Variation across sites due to regional or administrative structural differences
- Noncomparability of the Head Start and no–Head Start groups, with a range of approaches used to compensate for initial group differences (e.g., less-than-perfect attempts to match the groups, the use of waiting lists to make up the comparison groups)

A critique of the Head Start Synthesis study was provided by Gamble and Zigler (1989). They raised concerns that the Synthesis study had included studies having serious methodological limitations. They objected furthermore to the study's singular focus on outcomes versus lack of outcomes. They argued that more attention should have been focused on what the various studies revealed about the quantitative and qualitative assessment of program activities from site to site. A better understanding of the range of factors that contribute to Head Start program quality and outcomes was thus identified by Gamble and Zigler, as well as by others, as an important research direction for future research. Similar recommendations were again made in *Beyond the Blueprint: Directions for Research on Head Start's Families* (Phillips & Cabrera, 1996). This challenge has only recently been addressed in coordinated research at a national level with the creation of the Head Start Quality Research Centers (U.S. General Accounting Office, 1997).

The 1997 U.S. General Accounting Office report

Out of the more than 200 research reports examined for the GAO report (out of 600 manuscripts initially identified), the GAO analysis selected 22 studies that met all the criteria for inclusion. The 22 were studies that (1) involved Head Start participation in 1976 or later, (2) compared Head Start outcomes with those of comparison or normative groups, and (3) reported tests of statistical significance. Of the 22 studies meeting these three criteria, 16 compared Head Start with a no-preschool group, 3 analyzed gains by comparing Head Start group findings with normative test data, and 3 compared Head Start study participants with comparable children who attended other types of preschool or day care. Only 1 of the 22 studies employed random assignment to Head Start and no–Head Start groups. The following examination of the 22 studies addresses size and representativeness of sample, timing of follow-up assessment, types of comparison groups, diversity of study groups, and range of assessment measures.

Sample size and representativeness, timing of follow-up. Many studies examined by the GAO had small samples, and even the larger samples tended to eventually become smaller owing to missing data or attrition. The GAO report stated that as of 1997, no representative national samples had been employed to examine Head Start outcomes. The sample sizes of the 22 studies can be classified as follows:

- Five studies had *large* sample sizes—over 300 children or families. In these 5 studies, *N* ranged from 5000 (in a study using the database of the National Longitudinal Survey of Youth study), to 817 (in the next largest study), to 316 (in the smallest of these large-sample studies).
- Eleven studies had *moderate* sample sizes—from 100 to 300 children or families. In these 11 studies, *N* ranged from 108 to 258.
- Six studies had *small* sample sizes—fewer than 100 children or families, with *N* ranging from 37 to 85. Having such low sample sizes was considered to be a major limitation of these studies.

The 22 studies focused on by the GAO report employed three basic types of study design in their assessments of children and families: pretest/posttest; posttest/immediate; and posttest/follow-up. Of all the studies,

- Six conducted pre- to posttest assessment—3 of these conducted assessments within or at the end of the year, and 3 conducted, in addition to that, follow-up assessments 1 to 4 years later.
- Five conducted only posttest assessments at the end of the preschool year.
- Eleven conducted posttest assessments, with one or more times of assessment—ranging from 1 year later up to grade 6.

Types of comparison groups. As indicated previously, only one of the selected studies employed random assignment to study groups. Thus, another potential limitation of the studies involved the comparability of their comparison groups and Head Start groups. Just how comparable the various comparison groups were to the Head Start samples was not examined in the GAO report. The report indicated that typically the studies had matched their Head Start and no–Head Start groups according to one or more background variables, usually *socioeconomic level*. The investigators in some of these studies with comparison groups applied statistical methods to adjust for any lack of comparability of the no–Head Start group and the Head Start group on a variable such as *socioeconomic level*.

Diversity of study groups. Few of the 22 studies were designed to examine the potential differential impact of Head Start for diverse racial/ethnic, socioeconomic, or regional groups. Nor is there a body of research on Head Start effects according to differences in program factors—quality of services, staff development, facilities—although several such studies are currently under way.

Range of assessment measures. The GAO report indicated that the body of research they had identified as useful for examining Head Start effects was inadequate for making conclusions about the program nationally with respect to school readiness or health-related services. Of the 22 studies they examined, 16 included one or more cognitive measures, 5 included health or nutrition outcomes, and 5 included family impact. (Some studies are counted more than once in this breakdown because they included more than one type of measure.) Cognitive measures were generally based on standardized achievement and intelligence test scores. Sixteen studies found immediate developmental gains for Head Start versus no–Head Start children. These studies used a range of academic achievement tests; cognitive, health, social, and other developmental assessments; and family-functioning assessments. However, 6 studies reported either no significant Head Start effects or inconclusive findings when Head Start children and families were compared with no–Head Start groups on any of a range of achievement tests, developmental assessments, or other measures.

Of the 11 studies that included either longitudinal or follow-up assessments 2 years or more after Head Start, 5 found significant gains for Head Start children on some outcome measure—

either test scores or incidence of grade retention—when they compared them with a comparison group or national normative data; 6 of the 11 studies found no significant gains for Head Start children.

Major studies included in the GAO report. One of the studies examined in the GAO report was conducted by Currie and Thomas (1995), who employed analyses from a national sample of data from the National Longitudinal Survey of Youth (NLSY) and the National Longitudinal Survey's Child-Mother (NLSCM) file. They focused on several tests of cognitive development and achievement (Peabody Picture Vocabulary Test [PPVT]), on grade retention, and on other outcomes. Like investigators in several other studies included in the GAO review, Currie and Thomas focused their analyses primarily on a comparison of siblings, where one had attended Head Start and the other had not attended any preschool. Gains on *PPVT test scores* were found for Head Start children in the early elementary grades; however, these early gains, for the African American children in the study, were found to diminish by age 10; for White children, the early gains persisted into their adolescent years. Investigators also found less *grade retention* for White Head Start children versus their no–Head Start siblings, but they found no such difference between African American Head Start children and their siblings.

Of the health and nutrition studies in the GAO report, 2 studies found gains for Head Start children on health variables (e.g., immunizations, health exams). The Fosburg and Brown (1984) study of health and nutrition employed random assignment of study participants to Head Start and no–Head Start groups. They found that Head Start children were more likely to receive *medical* and *dental examinations* as well as *speech* and *vision examinations.* Further support for Head Start's contributions to children's health was found by the Hale, Seitz, and Zigler (1990) study, which employed a posttest design with a Head Start group, a Head Start–eligible group who were on a "waiting list" for Head Start, and a group containing middle-class nursery school children. The Head Start children had received more *age-appropriate health screenings* and *immunizations* and more *dental examinations* than the comparison groups had.

The central purpose of the GAO examination appears to have been to determine whether or not there was or will be a national-level evaluation of the impact of Head Start that can be applied to today's programs. The GAO review thus excluded studies prior to 1976. This meant excluding the Head Start Planned Variation research (reviewed earlier in this chapter) as well as some other earlier studies. One drawback of this decision is that it also limited the GAO's selection of Head Start studies to just 22 out of the 200 studies they had selected from an even larger pool of studies. This severely narrowed the likelihood of predicting directions or patterns of findings that may be expected regarding Head Start's potential impact. It also narrowed the possibility of discovering methodological lessons from the overall Head Start knowledge base.

Three Examinations of Research on Demonstration Programs

Several major reports have reviewed the research on what are often referred to as high-quality programs—Head Start demonstration programs or other kinds of demonstration programs for economically disadvantaged preschoolers and families (Barnett, 1995; Brown, 1978; Consortium for Longitudinal Studies, 1983; Gomby, Larner, Stevenson, Lewit, & Behrman, 1995; Kennedy, 1978; Lazar & Darlington, 1982; Lee, Brooks-Gunn, Schnur, & Liaw, 1990; Powell & Roane, 1996; Ramey, Bryant, & Suarez, 1985; Schweinhart et al., 1993; Schweinhart & Weikart, 1997; Zigler & Styfco, 1994).

In this review of the Head Start knowledge base, we include studies of several early childhood demonstration interventions, even if they were not Head Start model programs, because of the relevance of the populations they focused on—low-income children and families. Their inclusion is appropriate because these studies, which are often well designed, demonstrate that processes in early education and care can link to positive long-term as well as immediate effects for children from economically impoverished backgrounds. Furthermore, many of these preschool intervention models readily became Head Start models (in Head Start Planned Variation or later). Many continue to this day to influence to some degree Head Start program components, including curriculum and inservice training of staff.

In the following pages, we focus on three major integrative examinations of research on early childhood demonstration intervention programs for low-income children: the Consortium for Longitudinal Studies, the Center for the Future of Children report, and the literature review of the High/Scope Perry Preschool Project.

The Consortium for Longitudinal Studies

The Consortium for Longitudinal Studies (Consortium for Longitudinal Studies, 1983; Darlington, 1991; Lazar & Darlington, 1982) was a collaborative follow-up of the participants in 11 individual studies of various early childhood intervention programs designed for low-income children and families. Two of these studies—Experimental Variation, Louisville, Kentucky (Miller & Bizzell, 1984) and the New Haven Follow-Through Study (Seitz, Apfel, Rosenbaum, & Zigler, 1983)—included Head Start groups. The early childhood programs included in the Consortium's follow-up were based on various child development models or on specific areas of child development—Piaget-based developmental approaches, early language stimulation, achievement motivation, parental participation (e.g., High/Scope's Perry Preschool study in Ypsilanti, Michigan; the Parent Education Program study in north central Florida; the Early Training Project in Tennessee). The programs included center-based, home-based, and combination center/home-based models. Children who had participated in these programs included 2- to 5-year-olds, and the duration of their participation ranged from 1 to 5 years. There were over 2,000 children in the follow-up study in 1976, when approximately 74% of the 11 studies' original participants were able to be located.

To examine the average effects of program participation across the various programs, the Consortium investigators conducted multisample secondary analyses and compared the program groups to no-preschool or comparison preschool groups. Because there were various study designs (both experimental and quasi-experimental), the Consortium's analyses on some outcome variables were conducted by pooling the data across those studies having similar study designs.

Overall, the data analyses found program effects for children on a number of major variables (see Lazar & Darlington, 1982). When compared with no-program children, program children were more likely to have higher *intelligence test scores* lasting into their early school years, were less likely to be placed in *special education* classes, and were less likely to be *retained in grade*. The statistical analyses found that immediately after preschool, the preschool programs' pooled group intelligence test score was significantly above that of the pooled no-preschool control/comparison group. This advantage was declining but still significant up to 3 or 4 years after the preschool program. Furthermore, the study analyses indicated that on average, the preschool program reduced the number of children placed in special education classes and reduced grade retention.

The Center for the Future of Children report

An issue of the Packard Foundation's journal entitled *Long-term Outcomes of Early Childhood Programs* (see Gomby & Larner, 1995) presented 12 articles that focused on the effects of early childhood programs, including some Head Start programs, that served children from low-income backgrounds. In this issue, Barnett reviewed the cognitive and educational effects found in 36 studies of both demonstration and large-scale public school programs, including some Head Start programs. This review focused specifically on studies investigating the long-term effects of programs for children from low-income families (e.g., the Early Training Project, the High/Scope Perry Preschool Project, the Philadelphia Project, the Syracuse University Family Development Research Project, the Abecedarian Project, the Yale Child Welfare Research Project, and others). As in other reviews of the research, in Barnett's review, the studies of various programs were examined with respect to program description, age of program participants, research design, initial and follow-up sample size, time of follow-up, and types of measures used. These studies were conducted from the 1960s through the 1980s, and in a few cases, up until the present.

Summarizing the review, Barnett (1995) suggested that many would propose that only the Abecedarian and the High/Scope Perry Preschool studies are sufficiently valid but that the other studies have enough validity to add to the weight of the evidence. According to his analysis, taken as a whole, the reviewed studies, along with the validity of their findings, indicate that early childhood programs can result in sizable effects in terms of children's success continuing into their school years. These effects include an increase in *intelligence* and *achievement test scores*, a decrease in *grade retention* and in *special education placement*, and an increase in *high school completion*.

The literature review of the High/Scope Perry Preschool Project

The most recent report of the High/Scope Perry Preschool Project (Schweinhart et al., 1993) divided the major investigations of early childhood education programs into two major categories—*intensive studies* and *extensive studies*. Schweinhart and colleagues described *intensive studies* as those that are single-site, include under 500 study participants, employ a true experimental design, and focus on the individual as the unit of analysis. They described *extensive studies* as those that are multisite or multiyear in selecting children, include over 500 study participants, employ a quasi-experimental design, and may be set up to focus beyond the individual as the unit of analysis. They then observed that intensive studies are often *strong on internal validity* but weak on generalizability, whereas extensive studies are often *strong on generalizability* but weak on internal validity; together, these two types of studies may complement each other to determine the likely degree or nature of a program's impact.

Using this intensive-extensive framework, Schweinhart and colleagues selected and categorized the major longitudinal studies of programs for young children living in poverty who were considered to be at risk of school failure. Along with the High/Scope Perry Preschool study, these 7 studies of preschool programs (including Head Start and other preschool models) were categorized as being *intensive:* the Abecedarian Project (Ramey, Bryant, Campbell, Sparling, & Wasik, 1988), the Early Training Project (Gray, Ramsey, & Klaus, 1982), the Harlem study (Palmer, 1983), the Milwaukee Project (Garber, 1988), the evaluation of the Mother-Child Home Program (Levenstein, O'Hara, & Madden, 1983), the evaluation of the Rome Head Start Program (Monroe & McDonald, 1981), and the Syracuse University Family Development Research Program (Lally, Mangione, & Honig, 1988). The longitudinal studies they classified as *extensive* were these 4: the evaluation of the Chicago Child-Parent Center Programs (Fuerst & Fuerst, 1993), the evaluation of Project HOPE (Gotts, 1989), the evaluation of the Houston PCDC (Johnson, 1988), and the evaluation of the New York Prekindergarten Program (Irvine, 1982). Specific features of the High/Scope Perry Preschool study and of these additional 11 studies are summarized next.

Across the programs studied, the *age of child at program entry* ranged from birth through age 5. Across programs, the *length of time enrolled in program* ranged from 1 year to 5 years (at most up

BEST COPY AVAILABLE

to the end of early childhood, e.g., from birth to age 5, from age 1 to age 3, from age 3 to age 8). The *age of study participant at most recent follow-up* can be described as follows: the total of 12 studies included 1 study in which the most recent follow-up was at approximately age 4, 1 study in which it was at ages 8 to 11, 1 study in which it was at age 9, 1 study in which it was at ages 9 to 13, and 1 study in which it was at age 13. There were 3 studies in which follow-up was between ages 14 and 17 and 4 studies in which the young people were followed into young adulthood (age 19 or older; this latter group includes the High/Scope Perry study that followed participants through age 27). The percentage of original study participants who were followed varied considerably across the 12 studies: 2 studies had low follow-up response rates (17% and 43%); all other studies had follow-up response rates that ranged from 71% to 88%. Among this latter group was the High/Scope Perry Preschool age-27 study, which had a response rate over 90%.

A review of the *key outcomes* from the intensive and extensive studies selected by Schweinhart and colleagues (1993) revealed that in general, a greater number of statistically significant positive outcomes were reported by the intensive studies—perhaps because they had higher internal validity and more manageable sample sizes than did the extensive studies. Across the 12 studies reviewed (including the High/Scope Perry Preschool study), the statistically significant positive program effects found at one or more ages were these: *improved intellectual performance* in 7 intensive studies and 1 extensive study, *less special-education placement* in 4 intensive studies and 1 extensive study, *less grade retention* in 1 intensive study and 2 extensive studies, and *increased high school graduation* in 1 intensive study and 2 extensive studies.

Summarizing the Major Syntheses

The Head Start Synthesis study

The research (76 studies) reviewed in the Head Start Synthesis study (McKey et al., 1985) reveals a pattern of immediate positive effects of Head Start in preparing children for school and lowering the potential for difficulties in the ensuing school years. In a comparison of Head Start and no–Head Start children, the Head Start children's immediate positive effects on achievement *and* intelligence test scores *did not give them any significant advantage on these specific variables beyond grade 3.*

The 1997 GAO report

The GAO report's inclusion of 22 was studies based on very stringent criteria. In 16 of the studies it reviewed, immediate effects were found for Head Start children versus no–Head Start children on one or more of a range of academic achievement tests; cognitive, health, social, and other developmental assessments; and family-functioning assessments. Six of the 22 studies reported either no significant Head Start effects or inconclusive findings. Eleven of the 22 studies conducted either longitudinal or follow-up assessments, and 5 of these studies found long-term advantages for Head Start children on some outcome such as test scores or incidence of grade-retention; 6 found no significant differences over time. Even in light of the obtained findings for early Head Start effects, the report concluded that "Findings from the individual studies we identified, however, could not be appropriately combined and generalized to estimate program impact at the national level" (U.S. General Accounting Office, 1997, p. 8).

The three examinations of research on demonstration programs

The Consortium for Longitudinal Studies (1983) cross-study analyses, the Barnett (1995) analysis, and the intensive-extensive study analysis proposed by Schweinhart and colleagues (1993) demonstrate the value of intensive studies that are small-scale but highly valid. This gives support to some of the major recommendations in the Beyond the Blueprint *report (Phillips & Cabrera, 1996), which has called for future Head Start research to include a strong focus on small-scale studies that would involve diverse regions, programs, and samples and that would examine program-quality effects.*

SOME SPECIFIC LONG-TERM STUDIES OF DEMONSTRATION EARLY CHILDHOOD PROGRAMS

The research reviews described thus far reported that the knowledge base on Head Start and on other early childhood interventions contains some evidence of encouraging immediate and long-lasting effects on participating children. Several of the specific early childhood studies already cited in the Barnett (1995) review and in the intensive-extensive study analysis by Schweinhart and colleagues (1993) are especially relevant to the LTBHS study and are therefore briefly described next.

The Abecedarian Project

The long-term intellectual and academic outcomes for children and families in the Abecedarian Project have been reported by Campbell and Ramey (1995) and others. The study had initially randomly assigned families to either an early childhood intervention group or a control group. The early childhood intervention, which was an infancy-through-kindergarten program, was full-day and full-year. After completing the early childhood phase of the study, the study participants in each of the original two groups were again randomly assigned to either a new treatment group or a new control group, resulting in four types of groups, including two groups who would receive and two groups who would not receive elementary school intervention through age 8. Somewhat like the Head Start comprehensive approach and like many Follow Through programs, the elementary school portion of the intervention also included a home/school resource teacher, special activities for the home, and social services assistance. The long-term effects of the overall intervention program (infancy through age 8) were examined when those study participants were age 15. There were lasting effects—the program study participants, compared with the no-program study participants, tested higher on *reading* and *mathematics achievement test scores* and had fewer *grade retentions* and fewer *assignments to special education*. In another analysis that included the variable of *arrest rates* up to age 21, Clarke and Campbell (1997) found no preschool program effect on this particular variable when comparing all those in the preschool intervention with those in the no-preschool control group, irrespective of the elementary school intervention.

The Evaluation of the Chicago Child-Parent Center Programs

Reynolds (1992) studied the Chicago Child-Parent Center (CPC) programs with a sample of 266 children in CPC preschool programs and 125 children in a matched comparison group, all of whom were tracked from kindergarten through third grade. The study data included *preschool enrollment, cognitive readiness, teachers' rating of social maturity, parent involvement, special education placement, school mobility, grade retention,* and *reading and mathematics achievement* at third grade. Using a latent-factor structural equation model, Reynolds found that while some outcomes (e.g., *cognitive readiness* for school) were directly affected by study participants' having attended preschool, other outcomes (e.g., *later achievement* and *grade retention*) were affected indirectly by mediating school and family factors (e.g., parent involvement), rather than directly by the preschool experience. These findings illustrate the importance of considering factors that may detract from, lessen, or facilitate the positive outcomes of preschool experience.

The results of Reynolds's analyses support the hypothesis that by third grade, school and family factors, as well as the child's cognitive readiness, mediate the effects of preschool on grade retention and achievement. He found effects of preschool to be stronger given these conditions: that the children had entered kindergarten at higher levels of cognitive readiness, that the teachers had positively rated children's progress after preschool, that the parents were involved in the children's education, that the children had stability in the home as well as in the school, that the children had not been retained in a grade. The implication of this research is that studies should take into account factors that may mediate the effects of Head Start or other early childhood programs.

The Syracuse University Family Development Research Program Longitudinal Follow-up Study

The long-term effects of the Syracuse University Family Development Research Program (FDRP) were examined by Lally and colleagues (1988). The FDRP provided comprehensive prenatal-through-early-school services to low-income children and their families. Parent involvement was a cornerstone of the program, which viewed the parent as the child's primary teacher and caregiver. Lally and colleagues conducted a 10-year follow-up study of the children who had completed the 5-year program. The study included 108 children who started the program, 82 of whom completed the full 5 years of the program; of the matched controls, 74 children remained in the sample up to age 5. Sixty-five (79%) of the 82 children who had completed the 5-year program and 54 (73%) of the 74 matched controls who remained in the study sample up to age 5 subsequently agreed to participate in the follow-up study. Data from school records, teacher ratings, family interviews, parent interviews, child interviews, and court records were collected.

Compared with the no-program group, the FDRP children were found to express more positive *feelings about themselves*, take a more active *approach to personal problems*, and *view school* more positively. Program girls, but not boys, tended to benefit in the domain of school functioning—ending up with higher *grade-point averages*, *school attendance*, and *teacher ratings*. The strongest program effects were found in the domain of social deviance and functioning in the community: Only 6% of the program children, compared with 22% of the control children, were as teens processed as *delinquency probation* cases. The *severity* of the delinquent offense and the degree of delinquency *chronicity*, as well as the *costs of handling the delinquency cases* were higher for the comparison group. The study thus illustrates the important contribution that high-quality early education and care can make to later positive school experience and to social adjustment in the teenage years.

The High/Scope Longitudinal Studies

The research questions and design of the LTBHS study are especially related to those of High/Scope's Perry Preschool study and Preschool Curriculum Comparison study, both longitudinal studies conducted by Schweinhart, Weikart, and colleagues (Schweinhart et al.,1993; Schweinhart & Weikart, 1997).

The High/Scope Perry Preschool study

The High/Scope Perry Preschool study employed an experimental group design, with random assignment of children to either a High/Scope preschool group or a no-preschool, control group. The purpose of the study was to determine whether a high-quality early childhood program is better than no program at all for young children living in poverty. The High/Scope educational program is largely based on the child development principles of Jean Piaget. It includes activities and interactions with adults and peers that support and foster children's active, constructive learning and development (see Hohmann & Weikart, 1995). Over the nearly 25 years of the longitudinal study, High/Scope researchers found that the program using the High/Scope curriculum improved children's *early childhood intellectual performance*, later *school achievement*, *high school graduation rate*, and *adult earnings;* cut their *lifetime arrest rate* in half; and returned $7.16 to taxpayers for every dollar invested in the program (Schweinhart et al., 1993).

The High/Scope Preschool Curriculum Comparison study

The High/Scope Preschool Curriculum Comparison study employed a three-curriculum-group design (a traditional Nursery School group, a Direct Instruction group, and a High/Scope group). By randomly assigning children to these groups, it sought to determine the relative effectiveness of these three early childhood curriculum models for economically disadvantaged young children. Thus, the Curriculum Comparison study examined whether the High/Scope curriculum used in the High/Scope Perry Preschool study was essential to producing that study's lifetime results—or if other curriculum models could effectively have been substituted. Just as the High/Scope Perry Preschool study was designed to assess the value of high-quality early childhood education, the Curriculum Comparison study was designed to assess how broadly "high-quality early childhood education" could be defined in terms of differing theoretical bases.

The High/Scope Preschool Curriculum Comparison study found clear indications that by the time study participants were 23 years of age, the High/Scope and traditional Nursery School programs had produced better long-term outcomes than the Direct Instruction program had. Children who participated in the High/Scope and traditional Nursery School programs, compared with those in the Direct Instruction program, spent fewer of their school years in *special education* for emotional disturbance or behavior problems. By age 23, they had engaged in less *criminal activity* (including having fewer felony arrests), had done more *volunteer work* in the community, had more indicators of *economic viability* (e.g. car ownership, checking accounts), and had more *stable households* (e.g., were living with a spouse).

On a few variables, the High/Scope and traditional Nursery School programs showed advantage over each other (e.g., more of the High/Scope group returned to finish high school, higher annual earnings were reported by the Nursery School group). By contrast, the Direct Instruction group, although showing positive effects on intelligence test scores shortly after the preschool program ended, showed no advantage over the other two groups on any long-term outcome. In all likelihood, the Direct Instruction group fared no better over time than a no-treatment control group would have, at least not on the assessed variables. Overall, the results of the High/Scope Curriculum Comparison study suggest that the content and approach of a preschool curriculum can make a long-term difference.

Summarizing the Specific Long-Term Studies of Demonstration Early Childhood Programs

The five specific long-term studies just described—the Abcedarian Project, the evaluation of the Chicago Child-Parent Center programs, the Syracuse University FDRP follow-up study, and the two High/Scope longitudinal studies—have shown that strong preschool programs can result in early positive effects that can also have lasting impact—improved educational achievement, fewer negative school outcomes, and better social adjustment continuing into young adulthood. More small- and large-scale studies like these five are needed to determine the characteristics of program quality that are linked to obtaining some lasting positive effects, even when children go on to elementary and secondary schools of less-than-optimal quality.

BUILDING ON THE KNOWLEDGE BASE—THE LONG-TERM BENEFITS OF HEAD START STUDY

Researchers of Head Start can learn much from the study methods as well as the study findings reviewed in this chapter. For one thing, it is clear that many factors determine and limit the possibilities for estimating the effects of participating in Head Start versus participating in an alternative early childhood program or no program at all. The accumulated research identifies the necessity for paying careful attention to such major methodological considerations as these: identifying the sample and selecting subjects, determining the long-term study design and methods, accounting for mediating factors such as schooling and family background, and minimizing sample-selection bias and other threats to validity. All of these issues are addressed in Chapter 7's summary of the LTBHS study and discussion of the implications of the research that exists to date. The same issues are brought up with respect to specific implications for the LTBHS study in Chapters 3, 5, and 6, and in the reports by the study's two data analysts (Appendixes D, E, and F).

The Long-Term Benefits of Head Start study has had to deal with many of the major methodological issues brought up in this chapter's review of the Head Start knowledge base. For example, the LTBHS study encountered the following challenges of a long-term follow-up study with a quasi-experimental design: identifying, locating, and recruiting over 600 study participants in two sites, 17 years after their early childhood years; determining study participants' original-neighborhood socioeconomic characteristics; and developing appropriate statistical analyses. These methodological issues and the decisions that were made to handle them are discussed and described in considerable detail throughout the remainder of this report. As already noted, many of the LTBHS study's methodological and statistical approaches should be instructive to researchers employing quasi-experimental, longitudinal, or large-scale designs.

We should point out here that the LTBHS study is largely a retrospective follow-up study, because the original family-background data on participants had to be collected 17 years after they were preschool-aged, during the follow-up interviews in young adulthood. Although most of the former Head Start participants followed up in the LTBHS study had been a part of the earlier, Head Start Planned Variation study (Smith, 1973), preprogram and immediate postprogram data on these HSPV subjects were not available for subsequent analyses. This is because, owing to the concerns about confidentiality, the HSPV data analysis contractors had been required to scramble HSPV study participants' identification numbers so that the data would no longer be identifiable in terms of individual subjects. Upon inspection of this early data set (at the National Archives), we determined that in addition to the data's nonidentifiability at the individual-subject level, there was a considerable degree of missing data and subject attrition. Furthermore, since the LTBHS study added a No–Head Start comparison group and some additional Head Start subjects to the original HSPV study design, no parallel early data had ever been collected for these additional study participants. However, we were able to collect the earlier, family-background data on all LTBHS participants at their follow-up interviews; we also collected their school records data from kindergarten through postsecondary levels, as applicable, and their adult criminal and public assistance records, as applicable.

The specific research questions addressed by the LTBHS study were these:

- As young adults, are persons born in poverty who attended Head Start better-off than persons born in poverty who did not attend Head Start?
- As young adults, are persons born in poverty who attended a Model Head Start program (in this case, a Head Start program using the High/Scope educational model) better-off than persons born in poverty who attended a Regular Head Start program?
- Does Head Start have differential effects on its participants according to gender, racial/ethnic membership, or socioeconomic status?

3

THE DESIGN AND METHODOLOGY OF THE LONG-TERM BENEFITS OF HEAD START STUDY

When planning the design and methodology of the Long-Term Benefits of Head Start (LTBHS) study, we were able to use as resources a number of published reports of other long-term studies. We found helpful information in the High/Scope Perry Preschool longitudinal study through age 19 (Berrueta-Clement et al., 1984), in a 40-year follow-up study of middle-aged adults who had been identified as mentally retarded (Ross, Begab, Dondis, Giampiccolo, & Meyers, 1985), in the 10-year follow-up of the Family Development Research Program from Syracuse University (Lally et al., 1988), in the 10-year follow-up of the Family Support Intervention (Seitz, Rosenbaum, & Apfel, 1985), and in the research from the Consortium for Longitudinal Studies (Lazar & Darlington, 1982). Several research publications from The University of Michigan's Institute for Social Research (ISR) suggested useful strategies for the LTBHS study's instrument development and for its interviewing and coding procedures (e.g., for a recent summary, see Hill, 1992; Lansing & Morgan, 1971).

This chapter first provides background on the LTBHS study sites and groups, which are related to the sites and groups that were used in the national Head Start Planned Variation (HSPV) study reviewed in Chapter 2. Next, this chapter describes the steps taken to construct a No–Head Start group and to identify, locate, and interview the former Head Start and No–Head Start children approximately 17 years after they had entered elementary school in their communities. The chapter's final section provides descriptive data on the final study sample, details the data collection process, and presents an overview of the approaches used for data analysis.

THE LTBHS STUDY GROUPS

The LTBHS study employed two sites, one in Florida and one in Colorado, and the two major study groups at each site were these:

- **Head Start group**—a group of young people who had participated in Head Start programs in 1970–71; at the Florida site these Head Start participants consisted of a subgroup who had participated in a Head Start program using an approach based on child development theory (the High/Scope preschool model) and another subgroup who had participated in a Head Start program using an approach with no particular theoretical basis. These two subgroups are called, respectively, the *Model Head Start group* and the *Regular Head Start group*. The Model and Regular Head Start groups each included rural as well as urban subjects.
- **No–Head Start group**—a comparison group of young people who had never participated in Head Start or in any other early childhood education program.

THE LTBHS STUDY SITES AND GROUPS

The communities, or sites, we selected for the LTBHS study—which are two of the same communities that had been used in the Head Start Planned Variation study—are in northwestern Florida and northern Colorado. Chapter 2 explained how the HSPV study had compared *regular* Head Start programs, which had no particular theoretical basis, with *model* Head Start programs, which were based on specific theories or emphasized particular aspects or areas of children's learning and development, as described by Smith's report (1973). All the models were directly or indirectly developed by scholars in universities or in other educational organizations. Some models gave major attention to the importance of parent involvement, others to the role of the community, and others to the content of the curriculum and the teachers' interactions with their students. The models' developers held in common the belief that early intervention supports the overall development and learning as well as the eventual academic progress of children from economically disadvantaged backgrounds. The models were all implemented with attention to the cultural differences of the participating families and communities.

By contrast, the HSPV study's regular Head Start programs followed only the general guidelines provided by the newly developed national Head Start program, with local Head Start staff determining curriculum and providing support for the cultural differences of participating families and communities.

As described in Chapter 2, in the HSPV study's Florida and Colorado study sites, the children who had participated in the High/Scope model Head Start program in 1970–71, when compared with children who had been in the regular Head Start programs that year, made statistically significant gains in intelligence test scores. Thus, the HSPV study findings prompted us to return to these sites to conduct the 17-year follow-up of the LTBHS study. The following provides a brief description of each major site and the LTBHS study groups. For more statistical detail on the sites, Chapter 5 presents and discusses the specific site characteristics, based on U.S. Bureau of the Census data (see Table 5.4).

The Florida Site

The Florida site includes two urban areas, each of which is bordered by a rural area. The site has a sizable tourist industry and takes in two major military bases. From 1970 to the present, within each of the urban areas, there have been middle-income neighborhoods adjacent to low-income neighborhoods. The Head Start population from the urban areas included predominantly low-income families of African American background; from the nearby rural areas, it included low-income African American as well as low-income White families.

In many of the HSPV study's sites throughout the country, most of the eligible Head Start children were enrolled in *model* programs (as opposed to *regular* programs), thus making it difficult for HSPV researchers to construct an appropriate Regular Head Start comparison group from these sites. Therefore the HSPV study design typically utilized additional nearby areas that had regular Head Start programs, even though these areas were often not completely comparable to those with the model Head Start programs.

Identifying an appropriate comparison group for their Model Head Start group in one of the urban areas in the northwest Florida site was especially difficult for HSPV researchers because nearly every Head Start–eligible child in that area had enrolled in a model Head Start program (High/Scope). As a result, the HSPV study in this Florida site selected comparison subjects from a regular Head Start program in a nearby urban area, even though the urban area was larger and may not have been completely comparable to the urban area of the model Head Start subjects.

As part of their Head Start group, the

LTBHS study researchers used the regular Head Start and model Head Start groups of the HSPV study's Florida site. We also added to these original HSPV study groups

- A Regular Head Start group drawn from a nearby rural area, to serve as a comparison group for the Florida site's rural Model Head Start group. (Members of this added, or new, group were not in the HSPV study but were served by the same Head Start agency as HSPV study participants.)

For the comparison No–Head Start group, the LTBHS study researchers also selected, at the Florida site

- A No–Head Start group, drawn from neighborhoods and elementary schools identical to or near those of participants in the Florida site's Head Start group. (In the smaller urban area—the urban location of the HSPV study's model Head Start program—only a few Head Start–eligible children had not attended Head Start and were thus available for enrollment in the LTBHS study's No–Head Start group.)

At the Florida site, nearly every one of the study participants who had participated in the HSPV study's model Head Start program (in one of the Florida site's urban areas) subsequently enrolled in elementary schools participating in the national Follow Through program (this was 39 students, 24% of the LTBHS study's Florida Head Start group). The Follow Through program in Florida was based on the High/Scope educational approach, which had been used in the HSPV study's model Head Start program. Model Head Start and regular Head Start children from the Florida site's other urban area and its two rural areas did not, however, go on to attend elementary schools that participated in Follow Through. Furthermore, of the LTBHS study's Florida site No–Head Start children, only 8 (3%) subsequently went to elementary schools with a Follow Through program.

The Colorado Site

The Colorado site consists of a small urban area that includes food and farming industries as well as a university. The surrounding rural area contains large farms and cattle ranches. At this site, the HSPV study's Head Start participants were, for the most part, from low-income Mexican American families (former migrant-worker families from the southwestern United States who had settled in this Colorado area), but the group also included children from low-income White families. At this site, as at the Florida site, the HSPV study had compared children who attended model Head Start programs with children who attended regular Head Start programs. However, owing to a lack of identifying documentation on the Colorado site's HSPV study group memberships, we could not distinguish between model and regular Head Start participants for the LTBHS follow-up study. (See Chapter 2, p. 26.) Thus, the Head Start children who were followed up in the LTBHS study at the Colorado site were the same ones who had participated in the HSPV study, but in this 17-year follow-up, they were treated simply as Head Start participants, without any differentiation as to their model or regular Head Start participation in the HSPV study.

As at the Florida site, for comparison with the Colorado site's Head Start group, a Colorado No–Head Start group was developed for the LTBHS study. These No–Head Start study participants were selected from backgrounds and neighborhoods that were highly similar to those of the Colorado Head Start group. However, the HSPV study's Head Start group had already taken in the majority of the Colorado site's available low-income Mexican American children, because these had been the children at that site who were found most at-risk economically. Therefore a No–Head Start group with a comparable proportion of low-income Mexican American children could not be constructed from any nearby area. This meant that the Colorado site's No–Head Start group to be used in the LTBHS study necessarily ended up with a lower proportion of Mexican American children and a higher proportion of White children as compared with the Colorado site's Head Start group.

At the Colorado site, the Head Start group contained study participants who subsequently attended an elementary school participating in the national Follow Through program (this was 50 students, 39% of the LTBHS study's Head Start group). Of the LTBHS study's No–Head Start children, only 18 (27%) subsequently went to elementary schools participating in Follow Through. As at the Florida site, the Follow Through elementary schools at the Colorado site used the High/Scope model.

THE STUDY DESIGN

Many evaluators of model versus regular Head Start and of the corresponding Follow Through elementary school programs (e.g., Emrick, Sorenson, & Stearns, 1973; Kennedy, 1978; Smith, 1973; Stebbins et al., 1977) have noted that the HSPV research was limited by the fact that at the HSPV study sites, the available regular Head Start study participants were not entirely comparable to the model Head Start study participants in socioeconomic background, locality, or other factors. As we have noted, the Florida and Colorado sites selected for the HSPV study typically had elementary schools participating in the Follow Through program, and these Follow Through schools used the same theory-based educational models as were used in the HSPV model Head Start programs. The limitations of both the HSPV study and the subsequent Follow Through program evaluations were discussed by Stebbins and colleagues (1977):

> Because of the commitment to deliver a service program to those most in need and because the founders of Follow Through initially assumed it would have a powerful and highly visible impact, they did not give high priority to aspects of the evaluation design that could have made the experiment more sensitive to the measurement of treatment effects. In particular, no attempt was made to assign children randomly to treatment and comparison groups. In most cases the Follow Through participants were selected from among the most disadvantaged children in the community. (pp. 11–12)

As explained with regard to selecting the LTBHS study groups at the Florida and Colorado sites, so that the LTBHS study could investigate the long-term effects of having been in a Head Start program of any type (regular or model), it was necessary to construct a No–Head Start group to compare with each site's Head Start group from the earlier HSPV study. This post-hoc, No–Head Start study group (which we sometimes refer to as the comparison group) consisted of young adults who had no Head Start experience but were originally from the same low-income communities as were the study's young adults who had participated in Head Start. A few study participants in both the Head Start and the No–Head Start groups reported having had some group family day care or other child care experience in a home or church (not an early childhood education program, however).

In constructing the No–Head Start comparison group 17 years after the HSPV study, we encountered limitations like those faced by the HSPV investigators in selecting their study participants. However, as discussed by Stebbins and colleagues (1977), these problems are not unusual in program evaluation research and should not preclude collecting useful data:

> In summary, the national Follow Through evaluation is subject to many of the problems inherent in field social research (i.e., nonrandom assignment, unclear definition of treatment, problems of assessing implementation, less than ideal instrumentation) and careful judgment must be used in drawing conclusions. Evaluation of these planned variations provides us with an opportunity to examine the educational strategies under real life conditions as opposed to contrived and tightly controlled laboratory conditions. (p. 13)

The LTBHS study design made possible a number of different kinds of useful long-term comparisons. The most important one was a comparison of a Head Start group with a No–Head Start group, with both groups made up of young adults drawn from the original HSPV study sites in Florida and Colorado. Further, the Florida site provided a sufficient number of study participants for comparing urban and rural study participants within and between study groups. Both sites also provided an opportunity to do long-term assessment of various Head Start versus No–Head Start low-income groups—African Americans, Hispanic Americans, and Whites. Finally, at the Florida site, it was possible to compare Model versus Regular Head Start programs, thus providing a long-term evaluation of one of the HSPV Head Start models (i.e., the High/Scope model).

POTENTIAL SOURCES OF STUDY BIAS

Since none of the LTBHS study groups from the two sites were selected on a random basis, implicit or explicit selection factors may have resulted in some study bias—in differences between the Head Start group and the No–Head Start group that were independent of the Head Start experience. In the following, we discuss the two sources of potential study bias that we needed to take into consideration in selecting study methodology and interpreting study findings. One source has to do with the process of Head Start recruitment; the other has to do with the process of collecting study data.

Potential Bias in the Head Start Recruitment Process

In planning the approach for selecting the No–Head Start (comparison) group for the LTBHS study, we decided to identify those adults who, as children, had not been recruited or had not enrolled in Head Start but were from the same or similar socioeconomic backgrounds, neighborhoods, and schools as the Head Start children at each of the study's two sites. Given what we know about Head Start's practice of recruiting children and families with the most need, we expected that any No–Head Start group we constructed would represent somewhat less disadvantaged households than did the Head Start groups. Love and Grover's (1987) nationally representative study of Head Start found that low family income was the first major criterion used by Head Start agencies in recruiting children and families. Other factors included disability of the child, age-appropriateness, welfare dependence, and family problems.

More specifically, the 1970–71 Head Start program staff reported to us that they had conducted their recruitment of students for Head Start in the site's most impoverished neighborhoods. (This was also noted in reports on the HSPV study [e.g., Smith, 1973].) The Head Start staff also indicated that their recruitment had typically continued until all the available slots were filled. In this recruitment process, they focused on the lowest income neighborhoods as they (a) went door to door, going on to the next house when no one answered; (b) recruited on neighborhood streets, near stores at random times of the day; and (c) responded to calls of inquiry that came from parents who had heard about the program. Thus, some of the LTBHS study's No–Head Start group may have come from somewhat higher income neighborhoods, where there had been less-intensive Head Start recruitment. (See Chapter 5 for further discussion of the relative poverty levels of the study groups across sites. Note especially Table 5.4.)

Even if the LTBHS study had successfully constructed a No–Head Start group highly comparable to the Head Start group in socioeconomic background, it would have been impossible to totally overcome self-selection and other kinds of potential bias. For example, compared with families of children not enrolled in Head Start, the families of Head Start children may have been more able or more motivated to enroll their children. Furthermore, in the final selection of Head Start children, program staff may have chosen children who had certain characteristics that distinguished them from the children they did not choose. For example, the staff may have tended to enroll the child whose parent had most actively sought enrollment. Or, the staff may have tended to enroll the child of a parent who was easier to contact because the parent was at home during the door-to-door recruitment; this could indicate either that the parent was very disadvantaged economically (e.g., qualified for public assistance, not employed) or that the parent was somewhat better-off economically (e.g., more adequately supported by an employed spouse and thus not employed).

Another potential source of bias is the existence of family characteristics that led to Head Start enrollment and that might be associated with broader characteristics having a negative or positive effect on the child's development. For example, some parents who enrolled their children in Head Start may have done so because they held strong opinions about the importance of providing educational opportunities. Other parents, lacking adequate resources or support systems, may have sought Head Start enrollment as a way of providing babysitting for their children. Sources of both positive and negative bias could also be involved in the reasons why some families elected *not* to enroll their children in Head Start. Some parents who did not enroll their children may have had extended families that provided a high degree of support for child care and other needs. Other families who did not enroll their children may have felt too overwhelmed by difficulties in their living situation, lacked motivation to promote their children's achievement, lacked awareness of the developmental needs of their children, or tended to distrust a public institution such as social services.

Overall, in a study that does not employ random assignment to study groups, there is no way to eliminate the possibility of selection bias. Therefore, the potential influence of selection bias underwent scrutiny in the study's data analyses.

Potential Bias in the Data Collection Procedures

In the process of conducting the LTBHS research, it was hard to avoid various additional sources of potential bias. Some of the Head Start agency staff, especially those current or former staff members from the early 1970s, *did* participate in the study on a short-term basis, helping us to locate the 1970–71 Head Start students; their help consisted primarily of merely providing and verifying addresses and telephone numbers, as detailed in a later section of this chapter. Overall, however, there was little or no direct contact between these Head Start staff and the prospective study participants. When all location leads had failed, a Head Start staff member in each locality agreed to help locate study participants by making door-to-door visits to inquire about the most recent address of families of the potential study participants (including those from both the Head Start group and the No–Head Start group).

To guard against any potential Head Start bias on the part of study respondents or interviewers, we employed a number of strategies to minimize any association of the study with Head Start per se: (1) We were not specific about the study's Head Start versus No–Head Start focus; instead we instructed the LTBHS study's coordinators, locators, and interviewers to present the study as a follow-up of a school district's former students who may have been in a variety of school and preschool programs (we had the school district's permission to characterize the study in this way). (2) We made sure that the local project coordinators and the interviewers we hired had never been employed by Head Start. In addition, the interviewers did not personally know the subjects they interviewed or the families of those persons. (3) For the most part, we did not inform interviewers about the group status (Head Start or No–Head Start) of those they were seeking to contact and interview. At the end of the interview, the interviewer asked the study participant a question about his/her participation in a range of past educational experiences, and Head Start participation was included as only one of these possible past experiences.

Since precautions were taken to minimize the sources of potential bias in data collection procedures, the most probable sources of bias in the LTBHS study were the likelihood that the Head Start groups at the two sites were more disadvantaged economically than the comparison groups and the likelihood that many of the Head Start family members were especially receptive to, or eager to have, either (1) the child care or (2) the educational experience that Head Start could provide for their child.

THE STUDY PROCEDURES

Selecting Potential Members of the Study's No–Head Start Group

To form a pool of potential members for a No–Head Start group at each of the two study sites, we used the three criteria listed in the insert at the bottom of this page.

Our first step was then to determine the 1970–71 home addresses of (1) the children who had participated that year in Head Start and (2) the children who, on the basis of their poverty location/school, were likely to have been eligible for but had not been identified as participants in that year of Head Start at any point. Based on consultation with the U.S. Bureau of the Census research division on poverty, we determined the highest poverty tracts and areas at and near the HSPV study sites, using 1970 census maps. From these maps and the census data, we were then able to identify specific tracts and areas where the 1970–71 Head Start children had lived and to determine cutoff points on streets where households were more advantaged economically. We also consulted with Head Start and elementary school staff from 1970–71 to ask them to identify the streets and neighborhoods and the elementary schools where children eligible for but not participating in Head Start were likely to have been at that time.

When the schools of the former Head Start children provided us with an insufficient number of potential No–Head Start group members, we used nearby neighborhood schools to provide us with additional potential No–Head Start group members. In towns and rural areas where the U.S. Bureau of the Census had not tracted (provided neighborhood-level statistical information) because of insufficient population density in 1970, we selected potential No–Head Start group members from the households that were closest to and on the same or neighboring streets as the former Head Start children. Potential No–Head Start group members who originated from the same schools as (rather than the neighboring schools of) the Head Start children were identified for the more intensive recruitment.

Criteria for Potential Members of the Study's No–Head Start Group at Each Site

Persons were judged to be appropriate for recruitment for the LTBHS study's No–Head Start (comparison) group if they met each of these criteria (indicating eligibility for 1970–71 Head Start):

- In 1971–72, they and the members of the study's Head Start group lived on the same streets or in the same U.S. Census Tracts.
- By December 1971, they and the members of the study's Head Start group had initially enrolled in kindergarten or first grade in the same or neighboring elementary schools.
- They had never attended Head Start or another early childhood educational program.

As discussed earlier, one Florida urban area presented LTBHS study researchers with a unique challenge because most of its Head Start–eligible children had been enrolled in the HSPV study's model Head Start program. As a result, only a few comparable potential No–Head Start group members were available within that urban area. To achieve a pool of comparable potential No–Head Start group members, we had to also identify potential No–Head Start group members from the same nearby urban area where the HSPV study's regular Head Start children had lived.

To ensure a sufficient number of study participants and to maintain the potential for a later final matching of the study groups, the identified potential No–Head Start group was larger than the identified potential Head Start group. This over-enrollment would allow a later final matching, including if necessary an elimination of those No–Head Start study participants whose interview data might reveal that in 1970–71 they had been significantly different from the Head Start study participants in socioeconomic background. Over-enrollment of the No–Head Start group also would allow us to eliminate any potential group members whom we later discovered to have attended some Head Start program outside of the study site. In a study site where the potential No–Head Start group members far exceeded the number needed for the No–Head Start group, we used random selection with stratification by gender and ethnic background to identify a somewhat smaller number for location and interviewing.

Locating, Recruiting, and Interviewing Potential Study Participants

The goal for locating and recruiting the identified potential participants in both study groups—Head Start and No–Head Start—was to achieve representativeness. In order to achieve this representativeness, the potential members of both groups were very actively pursued, regardless of where they currently lived. A related procedural rule was to locate and interview a roughly equal proportion of "hard-to-get" students from both groups. Because we located and interviewed potential study participants even if they no longer lived in the area of the study sites, any potential bias introduced by migration of Head Start students or comparison students was probably minimized. The following section describes the procedures we followed to achieve our location and recruitment objectives.

Selecting Local Project Coordinators, Locators, and Interviewers

It is important to note again that to minimize response bias, the local coordinators and interviewers of the LTBHS study were *neither* present *nor* past employees of Head Start. As indicated earlier, to facilitate the location of study participants, some current or previous employees of Head Start were hired on a short-term basis as locators, to help us find the addresses or telephone numbers of prospective study participants. All locators were advised to refer to the study as a follow-up of school district students who may have been enrolled in a variety of school and preschool programs. No details regarding the specific purposes of the follow-up study were made available to either locators or interviewers. Coordinators, locators, and interviewers at each site came from the same diversity of racial/ethnic backgrounds as did the study participants.

Local project coordinators

To accomplish the data collection, part-time local project coordinators were hired within each study site. Each coordinator had previous research or comparable experience. Some coordinators lived in or near the study participants' current or previous neighborhoods. The coordinators who did not live in these neighborhoods were experienced in working with people from many diverse backgrounds. All coordinators had considerable contacts in the community and were involved in one or more community networks, such as the local school system or human services department. The coordinators assisted the LTBHS study director (from the High/Scope Educational Research Foundation's research division) in the hiring, training, and direct supervision of the locators and interviewers. They also conducted a large share of the location work, which included mailing information and contacting potential study participants or family members on the telephone. Throughout the follow-up interview data collection (which took about 3 years, from late 1986 through early 1990), a total of four Florida and three Colorado coordinators were supervised by the LTBHS study director, who made frequent site visits, telephone calls, and mailings of specific instructional procedures.

Locators and interviewers

Locators came from diverse racial/ethnic backgrounds and were selected for their particular skills or contacts with schools, preschools, Head Start, human service agencies, and so on. Although a few of the locators also served as interviewers, many others (either because they were judged to be not appropriate as interviewers or because they were not interested in being interviewers) only worked at locating study participants.

The interviewers either had completed at least some college-level education or had considerable professional experience working with people. They needed to be able to master the content and format of the interview questions and interact with interviewees in a professional and natural manner. They were often assigned to interview study participants who matched them in ethnicity or gender. Interviewers included males and females from African American, Hispanic American, and White racial/ethnic backgrounds.

As mentioned previously, to protect confidentiality and minimize potential study bias, the interviewer assigned to interview a person did *not* previously know the person or the person's family. Along with the local study coordinator and the locators, many interviewers also made phone calls to help in locating potential study participants.

Locating and Contacting Potential Study Participants

By using a variety of methods, we were able to locate most of the potential study participants. These methods are described briefly here and more specifically in Appendix A. In the study's low-income communities, community members' wariness of researchers, of professionals associated with the schools, or of social service organizations sometimes had to be overcome. Although we used many different strategies to counteract this wariness, working with a community-based network of civic leaders, educators, and other professionals proved to be especially effective. Working with such a network also helped us to locate and gain the cooperation of the potential study participants. Thus, many kinds of help from the community—from individual schools, Head Start, churches, various other agencies, and key neighborhoods—were instrumental in locating and recruiting the study participants, as we go on to describe next.

The approach used to locate participants

Study site coordinators, locators, and interviewers (all hired by the High/Scope Foundation for the study) identified at each site the key community leaders and long-time neighborhood residents who could provide leads on the current addresses and telephone numbers of the families and young adults who used to live in the neighborhoods of the 1970–71 Head Start children. The entire network used in the location process included former Head Start and public school staff (e.g., program directors, teachers, teacher aides, parent coordinators, Follow Through project staff), parents, former students, and key community leaders from the early 1970s. The local school districts involved gave their permission to be identified by LTBHS research workers as "cooperating with the follow-up study of their former students."

To gather lists, notes, and class photographs of the former school-system students and Head Start children, the LTBHS study director or the

BEST COPY AVAILABLE

local study site coordinator held group meetings and contacted individuals in the schools, Head Start, and the community. As mentioned previously, some current or previous Head Start program staff members agreed to visit the potential study participants' original neighborhoods to check on addresses, telephone numbers, or current whereabouts of families and former students.

We also did manual or computer searches of additional rosters, lists, and databases—high school student directories, state driver license registers, voter registration lists, U.S. Post Office change-of-address records, military base listings, prison lists, and college enrollment records.

The usual location process was for the project coordinators, locators, and interviewers to contact potential study participants directly or to contact their families or other community members (teachers, social workers, church members) to inquire about the former students and their parents or to verify addresses and telephone numbers. All information that was obtained concerning the whereabouts of the potential study participants (even the somewhat vague information) was given to coordinators.

The location obstacles encountered and overcome

Locating potential study participants 15 to 17 years after they were children of preschool age presented many obstacles. Many of the available addresses and telephone numbers were from the potential study participants' elementary school or high school years. These addresses often proved to be so out-of-date that they left no leads to pursue. School personnel were consulted to check on cases without leads, but they were in no position to put in the time and effort needed to trace their former students.

Sometimes, even after a potential study participant was initially contacted, only a few days later—perhaps by the time of the interview scheduling or appointment—that person had moved or had a telephone disconnected. Such quick changes of address and other similar location problems were due in part to the general mobility of young adults of this age (e.g., Lally et al., 1988). Furthermore, as often cited in other research studies conducted in low-income communities, potential study participants frequently (1) had their telephones disconnected because they could not pay the bills; (2) were evicted from their residences; (3) moved in with others when they ran out of money; or (4) were avoiding bill collectors, welfare department case workers, the police, or personnel from the school system. Others simply did not have a telephone. The persons hardest to find were those who were not in close contact with the family or were living somewhere on only a temporary basis. Although some potential study participants had moved away with their families during their school years, many times someone in their old neighborhood was able to tell us where the family now lived, and that information enabled us to locate the participant.

The project coordinators, locators, and interviewers, in consultation with the study director, clearly had to develop many approaches to overcome the problems of locating potential study participants. As specified later in this chapter, most of the former students at each site were found to be still living within the state, although most often in a different neighborhood or part of town, in a different nearby town, or in a different urban area within the state. Some were away at college, in the military (or married to someone in the military), or had moved permanently to another state.

The procedures used to contact participants

Once any potential study participant had been located, the local study coordinator or some assigned interviewer would contact that person or a family member, verify the potential study participant's identity (e.g., check on the birth date, names of parents, schools attended), explain the study, answer any questions, and explain that information about the study would subsequently be mailed out. The interview with the study participant was either scheduled at the time of this initial contact, or it was arranged that it would be scheduled by an interviewer or coordinator at a later time.

After this contact, information about the study in the form of a letter on High/Scope Foundation stationery was mailed to each potential study participant (or to a family member). These letters provided an advance verification and description of the study, usually prior to the local study coordinator's or interviewer's call to schedule an interview. Several kinds of letters were used to explain the study or elicit the interest of potential study participants (or of a family member, when a study participant's whereabouts had not yet been revealed or the participant was difficult to contact directly). Although most of the letters were brief, simple, and readable, we sometimes provided letters that explained more of High/Scope's background and experience and provided more information on the research. In these letters, care was taken to avoid any statements about the study's specific Head Start focus, which might later influence interviewee responses. Letters were also mailed to give verification for uncertain parents or students, to confirm a scheduled interview, or to indicate regret when a subject for some reason did not show up for his/her interview.

When parents had been located and the potential subject had not, parents who refused to provide their young-adult son's or daughter's address or telephone number often did instead agree to forward to them a letter explaining the study and containing an enclosed stamped envelope for returning the contact information. Alternatively, some parents agreed to ask their son or daughter to call us. About half of the time these third-party strategies resulted in the information getting to the potential study participant. Along with our explanatory letter, we enclosed a postcard or a form with spaces for the recipient to supply telephone numbers and best times for us to call. For study participants no longer living within the study site area, a postcard was provided to encourage them to notify us prior to their coming home for a family visit so they could be interviewed then. Interviews were sometimes successfully prearranged by the parents to coincide with a son's or daughter's visit to them from out of town. Telephone calls were also frequently made to confirm scheduled interviews and to check or reschedule if an interviewee did not show up or canceled the interview.

After a study participant was interviewed, particularly if it was a study participant who seemed to know many of his or her former fellow-students, the coordinator, and sometimes the interviewer, would show a list of not-yet-located potential study participants to the interviewee to inquire whether he/she knew the whereabouts of them or their families. If a particular interviewee's mother had been active in the school or community, she was also often a good resource for obtaining leads on the location of other prospective interviewees.

Interviewing Potential Study Participants

Most potential study participants we contacted were initially very interested in the research project and very responsive to participating. However, the interviews with them were sometimes difficult to arrange, as we now describe.

Study participants who were difficult to schedule

Most study participants were interviewed as scheduled, but a sizable number had to be rescheduled several times (because of cancellations or not showing up) before they were actually interviewed. Some study participants were interviewed only after many attempts, after a period of one to several months had passed. In an effort to schedule an interview, coordinators and interviewers would have to periodically recontact some of the persons who were difficult to schedule. Often, eventually catching a study participant at a time when he or she was not busy would result in the interview at last being successfully scheduled and carried out.

Another approach was to have the coordinators identify those interviewers who were readily available to conduct interviews. Whenever a potential study participant was successfully contacted and found willing to be interviewed, the coordinator would then attempt to set up an interview to be done that same day or the next, before the person got too busy or changed his/her mind. As mentioned, the major problem was that many potential study participants agreed to be interviewed but then would repeatedly not show up for the appointment, without ever verbally declining to be interviewed. After several periodic appeals for their participation, an estimated 5% to 10% of the potential study participants whom we directly contacted definitely or essentially declined to participate.

Study participants living outside the study site

Most of the LTBHS study interviews were conducted within the area of one of the study sites (81% of the 627 persons interviewed lived within the state of either the Florida site or the Colorado site, and 15% lived outside of the state; the addresses of the remaining 4% were not accurately recorded from the interviews). The 92 study participants who lived outside the study-site states resided in the following regions:

- *U.S. Northeast* (New England and Middle Atlantic states): 6 study participants
- *U.S. Midwest* (East North Central and West North Central states): 8 study participants
- *U.S. South* (South Atlantic, East and West South Central states): 45 study participants
- *U.S. West* (Mountain and Pacific Coast states): 28 study participants
- *Canada, Germany, or Korea:* 5 study participants

Some of these 92 study participants living out-of-site were interviewed when they visited their parents' home within the original study site. To meet with others, the interviewers traveled to other parts of Colorado and Florida and to several states throughout the South and West—Alabama, Arizona, California, Georgia, Louisiana, North Carolina, Texas, and Wyoming. We also hired and trained several interviewers who lived in various states and one who lived on a military base in Germany to conduct interviews of study participants who lived in their areas. The local study coordinators made the arrangements for the interviewers' out-of-town trips and sometimes had to do the interviewing themselves. For safety reasons, the coordinator or interviewer always had one other adult, either another study staff person or a spouse, accompanying them on out-of-site interviews. Only a small number of the 627 study participants (12, or 2%) were not interviewed in person. Eleven of these were interviewed by telephone, and 1 person filled out the interview and mailed it to us from overseas.

The last-interviewed study participants

As found in other long-term studies (e.g., Lally et al., 1988), the study participants who were the last 50–60 interviewed in this LTBHS study tended to be living in very unusual or difficult circumstances (i.e., avoiding bill collectors, using drugs, or avoiding an abusive spouse). These kinds of circumstances were also found among the earlier-interviewed study participants, but to a lesser extent. In a few (11) cases, although the potential study participant had been successfully contacted and found willing to be interviewed, throughout many attempts to hold the interview, a continuing special circumstance prevented the interview from actually taking place (e.g., the individual was in the hospital or recuperating from an illness; had emotional or cognitive limitations that would have made the interview process overly burdensome; or had a highly unusual work schedule, such as in migrant farming). For these 11 study participants, a parent agreed to provide the key interview information for the background and current-status questions (attitudinal questions were eliminated). In several of these cases, the study participant was able to be present at the interview, along with the parents, but had difficulty responding to the questions.

A few study participants were found to be in prison, and so we were able to locate them through the directory assistance of the state and local prison systems. Arranging the interview within a prison system was sometimes difficult or time-consuming, however. Nevertheless, these study participants were very receptive to being interviewed. We did not conduct prison searches outside of Florida and Colorado except in a few cases, when we were given a lead indicating that a potential study participant was in prison in a bordering state.

Interviewee compensation

The interviewees received a $25 check after the interview, to compensate them for their time. This compensation appeared to be a positive incentive for study participants of both the Head Start group and the No–Head Start group.

RESPONSE RATES OF THE POTENTIAL STUDY PARTICIPANTS

The interviewing effort resulted in a highly representative sample of each identified potential study group. (See Table 3.1.) Overall, the interviewed study participants were 77% of the identified potential study participants. Breaking results down by sites, the interviewed were 80% of those identified at the Florida site and 71% of those identified at the Colorado site. Table 3.1 shows that across the different subgroups at the Florida site (Head Start vs. No–Head Start, Model Head Start vs. Regular Head Start) comparable percentages of identified potential participants were interviewed; the Colorado Head Start and No–Head Start subgroups, however, had somewhat less-comparable percentages (80% vs. 63%) interviewed. The identified potential study participants who were *not* interviewed were those who could not be located, who refused to be interviewed, or who did not appear at any of several appointments set for their interview. The 627 young adults in the interviewed study sample were on average approximately 22 years of age and comprised 323 males (52%) and 304 females (49%); 338 African Americans (54%), 118 Hispanic Americans (19%), and 171 Whites/others (27%).

THE FINAL STUDY SAMPLE

A final adjustment of the interviewed sample was necessary. Although lists of all the 1970–71 Head Start children had been available at the Florida

Table 3.1: Potential and Interviewed Study Participants by Site and Group

	Identified Group	Interviewed Group	
Site/Group	*N*	*N*	% of Identified
Florida site total	**530**	**426**	**80.4**
Head Start	206	162	78.6
Model	93	75	80.6
Regular	113	87	77.1
No–Head Start	324	264	81.5
Colorado site total	**283**	**201**	**71.0**
Head Start	132	106	80.3
No–Head Start	151	95	62.9
Overall Head Start total	**338**	**268**	**79.3**
Overall No-Head Start total	**475**	**359**	**75.6**

site, children's participation at the Colorado site in 1970–71 Head Start had to be verified by their listing on one or more types of records (e.g., health records forwarded from Head Start to the elementary school, Head Start parent-participation forms kept by the human services department, Follow Through data files indicating children's earlier Head Start participation) or by two or more former Head Start staff and a parent. As part of each study participant's interview, the study participant was asked to indicate which of a list of various educational experiences he or she had participated in, and this list included Head Start as one possible choice. The data from this interview, together with the other kinds of verifying data, provided us with additional sources for affirming the Head Start status of each study participant at the Colorado site. Furthermore, if there was some indication from a record or from a Head Start staff member that a person from our pool of potential No–Head Start study participants may have participated in Head Start, the individual was excluded from the final No–Head Start group.

In the end, we had to eliminate 2 persons from the potential study participants we interviewed in Colorado, because whether or not they had participated in Head Start could not be sufficiently verified by any of the means just described. Also, we eliminated 1 of the interviewees from the Colorado site and 2 of the interviewees from the Florida site from our final study sample because their verified birth dates were discovered to be before 1964 or after 1967. It was not until after the interviews that we learned that a few children on Florida's 1970–71 Head Start lists had been outside the appropriate age-range but had nevertheless participated in Head Start because they were children of Head Start staff members. Overall, 97.1% of the final study sample were born in 1965–66.

Thus, after we had eliminated 5 of our interviewees (3 in Colorado, 2 in Florida), the final study sample for the Long-term Benefits of Head Start study consisted of 622 young adults who had been Head Start children or Head Start–eligible children in 1970–71 in Colorado or Florida. These were the study participants to be included in the final data analyses. Table 3.2 on the facing page shows the final LTBHS study sample by group, site, gender, and race/ethnicity.

The data analyses, reported in more detail in Chapter 5 and in the reports by the independent data analysts (Appendixes D, E, and F), provide an initial examination of these final study groups. The analyses showed in both Florida and Colorado that the Head Start and No–Head Start groups did not differ significantly by *gender*, that is, in their proportions of males and females. In Florida, the analyses showed there was also no significant difference between the two groups on *racial/ethnic membership*. At the Colorado site, however, there was a significant difference between groups on *racial/ethnic membership:* There was a greater proportion of Hispanic Americans in the Head Start than in the No–Head Start group. At the Florida site, the HSPV Model Head Start group (using the High/Scope model) and the Regular Head Start group did not differ significantly by *gender* or by *racial/ethnic membership*.

The analyses of the interview data from each site also showed that compared with members of the No–Head Start group, the members of the Head Start group came from more impoverished families. The Head Start group's interview data indicated that as children, these subjects had more single-parent households, larger families, fathers with lower educational levels, and overall lower socioeconomic status. Thus, on the one hand, Head Start staff had effectively recruited from the most impoverished families, as they are required to do. On the other hand, the resulting socioeconomic differences between our program and comparison groups would need to be taken into account in the analyses conducted to estimate the long-term benefits of Head Start. The socioeconomic backgrounds of the study groups are discussed extensively in Chapters 5 and 6 and in the two data analysts' reports.

Table 3.2: Final Study Sample by Group, Site, Gender, and Race/Ethnicity

		Head Start			No–Head Start		
Site & Race/Ethnicity	Total	Male	Female	All	Male	Female	All
Florida	**424**	**81**	**79**	**160**	**131**	**133**	**264**
African American	337	66 (81.5%)	59 (74.7%)	125 (78.1%)	103 (78.6%)	109 (82.0%)	212 (80.3%)
White/other	87	15 (18.5%)	20 (25.3%)	35 (21.9%)	28 (21.4%)	24 (18.0%)	52 (19.7%)
Colorado	**198**	**72**	**58**	**130**	**35**	**33**	**68**
Hispanic American	117	54 (75.0%)	41 (70.7%)	95 (73.1%)	11 (31.4%)	11 (33.3%)	22 (32.4%)
White/other	81	18 (25.0%)	17 (29.3%)	35 (26.9%)	24 (68.6%)	22 (66.7%)	46 (67.6%)
Total	**622**	**153**	**137**	**290**	**166**	**166**	**332**

Note. At the Florida site, the African American group included 329 participants who identified themselves as African American plus those study participants who identified themselves as Hispanic American (*n* = 4) or African American and White (*n* = 4). At the Colorado site, the Hispanic American group included those study participants who identified themselves as Hispanic American (*n* = 112) or Hispanic American and White (*n* = 5). Across the Florida and Colorado sites, the White/other group included 132 study participants who identified themselves as White plus those study participants who identified themselves as Asian American or Pacific Islander (*n* = 2), American Indian (*n* = 8), or combinations of other ethnic groups (*n* = 26).

COLLECTION OF STUDY DATA

This section describes the development of the LTBHS study's interview instrument, the types of records from which we collected data, and the data collection procedures. At the time of the interview, study participants also completed a literacy-assessment instrument, the Adult APL Survey (American College Testing Program, 1976). The LTBHS study's records search collected data from the U.S. Bureau of the Census, school records, criminal records, and public assistance records.

The Interview and Adult APL Survey

The interview instrument used with the young-adult subjects in the LTBHS study was largely developed from the one initially used in the age-19 study of the High/Scope Perry Preschool Project (Berrueta-Clement et al., 1984; Schweinhart et al., 1993). Likewise, the Adult APL Survey used in the LTBHS study was an updated adaptation of the version used in High/Scope's age-19 study.

The interview instrument

The interview instrument of the LTBHS study presented study participants with a fixed-format sequence of closed questions (i.e., yes/no questions, nominal assorted-choice items, and rating-scale items) and some open-ended questions (i.e., questions with no specified range of answers). The final form of the instrument was developed through an empirical, pilot-test process. To evaluate the comprehensibility of the interview questions and items, the instrument was first administered to approximately 50 young-adult non-study participants from the study site comunities. Based on the pilot-test participants' difficulties in understanding the questions and their logical sequence, we then refined the interview questions. The revised instrument was again piloted with additional non-study-participants from each study-site community. The interview's content, language, format, and question sequence were then finalized. On average, the LTBHS study interview took approximately 1 to 1½ hours to complete.

The study interview was designed so the interviewer would ask each question as printed on the instrument, rephrasing questions when necessary to ensure the interviewee's comprehension. The interview questions covered all the following topics having to do with the study participant's past and current life experiences and situations:

1. *Family background* from childhood to the present time, including parents' education and employment, household composition, number of family and household members, and the study participant's perceptions of parental expectations
2. *Current family status,* including the study participant's own marital status, parental status, and household composition
3. *Highest education completed,* an educational profile of the study participant, including the years of schooling and higher education completed by the study participant, the areas of study pursued, the courses taken, and other educational experiences and achievements
4. *Employment history,* an employment profile of the study participant, including the years or periods of time employed, the types of jobs, and other employment experiences or accomplishments
5. *Community activity,* including the study participant's participation in any community service or in any delinquency or crime
6. *Sources of economic support,* including the study participant's wages earned and the economic support received from family, government, and others
7. *Attitudes* held by the study participant toward self, work, education, and the future
8. *Sources of influence* perceived by the study participant, including role models or influential experiences in family, school, or community life
9. *Future goals and plans* of the study participant, including any educational and personal goals and plans

As indicated earlier, the interview instrument had been developed and employed with adolescents and young adults over the previous 20 years in ongoing studies investigating the long-term effectiveness of the High/Scope Educational Research Foundation's educational programs. The format, the content and focus of questions, and the survey methods of the interview instrument are based on reviews of the relevant theoretical and empirical research; findings obtained from ongoing research at High/Scope (Berreuta-Clement et al., 1984; Oden, Kelly, Ma, & Weikart, 1992; Schweinhart et al., 1993); consultation with researchers from the University of Michigan's Institute for Social Research; and the pilot work conducted specifically for each study.

The Adult APL Survey

The Adult APL Survey, developed by researchers at the University of Texas in Austin in 1976 and published by the American College Testing Program, is a general-literacy assessment instrument. Although the 1976 version of the instrument is no longer published, we obtained permission from the test developers to revise the wording of some questions of the 1976 version to make it less dated (e.g., we updated the costs of advertised products, the name of the current President of the United States). The Adult APL

Survey is a 40-item, multiple-choice instrument covering the topics of using community resources, gaining employment, being a consumer, and understanding laws. The questions cover a range of literacy tasks that require the study participant to solve simple arithmetic problems, write checks, follow maps and schedules, read employment announcements, and make sense of advertisements.

The APL Survey took about 20 minutes to administer and was usually administered prior to the LTBHS study interview. However, when an interviewer judged that the study participant appeared to be somewhat apprehensive about being interviewed, the APL survey was administered after, rather than before, the interview.

Interviewer training

The interviewers (those who administered both the APL survey and the interview instrument) were hired and trained by the LTBHS study director, with support and further supervision by the local-site project coordinators. The LTBHS study director at the High/Scope Research Foundation conducted most of the training sessions; a local project coordinator individually trained a few interviewers who could not attend these training sessions. The interviewer's training session included receiving instruction from the director or coordinator, reading a manual, witnessing interview demonstrations, and role-playing the interview. Next, the interviewer conducted several pilot interviews (with non-study-participants). These pilot interviews were tape-recorded for review by the study director, who then provided feedback. If the tapes and subsequent feedback indicated that a potential interviewer was unlikely to be an effective interviewer, that person was not hired for the study. Some of the interviewers who were hired actually ended up conducting only a small number of interviews, because they either found it too difficult to schedule interviews or had several interviewees who did not show up for the interview. As a result, a small number of interviewers conducted most of the interviews for each site.

Interview and Adult APL Survey data collection

Most interviews were conducted in person, with only a few conducted by telephone. During the interview, the interviewer wrote down the study participant's responses. When the study participant gave permission, the interviewer tape-recorded the responses to the open-ended questions (e.g., the question about being influenced by role models).

To facilitate data collection, the interviews were conducted in a variety of places. As explained earlier, the interviewers traveled when necessary to other cities and states, military bases, colleges and universities, hospitals, and prisons to conduct interviews with study participants no longer at the study sites. The interviews—both at the sites and away from the sites—took place in a variety of localities—in the study participant's home; occasionally in his/her workplace; or in a library, restaurant, hotel lobby, or local school building. It is important to note that to avoid potential bias, interviews did *not* take place at Head Start centers.

The interview was presented to study participants as a follow-up of former students from the area schools to learn about their lives now that they were adults. As explained earlier, the interviewer most often first learned of a study participant's Head Start participation at the end of the interview, when the respondent was asked about attending any of a list of specific school or community programs, with Head Start included on the list (see pp. 32–33.)

When an interview was completed by an interviewer, the completed instrument was submitted to the local project coordinator, who was trained to check with the interviewer to clarify hard-to-read handwriting or any incomplete or incomprehensible sentences. The completed interviews were then submitted to the study director, who would frequently review them for clarity and completeness.

BEST COPY AVAILABLE

Because of the many obstacles we encountered in locating and recruiting study participants, the data collection by interview took somewhat more than 3 years—it began in late 1986 and lasted until early 1990; we collected 74% of the interview data in 1987, 1988, and 1989. For this reason, the study participant's age at the time of interview had to be included as a major variable in the data analyses, as reported in Chapters 5 and 6 and in the two data analysts' reports.

The Records Search

In the LTBHS study, it was important to search certain records to obtain external, independent data to supplement the self-report data derived from the interviews. Such information is also potentially useful in conducting a cost-benefit analysis like the one conducted using the records data from the High/Scope Perry Preschool project (Barnett, 1996). That analysis indicated that attending preschool was cost-effective in the long run owing to reduced public costs of crime and welfare assistance.

In the LTBHS study, so we could have a common frame of reference for the study participants' records data (including data from the current year and from prior years), the information from school records, public assistance records, and criminal records was collected within specified time frames, as discussed in each of the following sections.

Obtaining access to records data

After being interviewed, study participants gave written consent for us to collect records data. We also secured the necessary consent from the schools and agencies involved in the records search. Schools and agencies also provided some staff support for conducting the search of their records. Criminal records were generally attainable, because such records are open to public access. Determining whether a person had ever received public assistance was also possible, although reviewing the public assistance records required obtaining permission from the individual study participant.

Collecting the records data on participants such a long time after 1970–71 presented many challenges. There were often delays involved in gaining clearance to proceed with our records work within the respective institutions. Study staff often had to conduct manual searches in out-of date, unused file-storage systems to locate and retrieve study participants' records; in some cases, it was possible to work with newly installed computer systems to collect data from the more recent records. It was often necessary to request school and agency staff time to assist us in searching and collecting records, and we also needed to obtain agency clearance for our staff to record the information.

Educational records

Data collection from school records took place in the districts where study participants had attended elementary and secondary school. This included local schools and sometimes also schools in other localities (for study participants who had relocated). School districts, including those not within the study sites, provided the study participants' school records information, with technical support from the study staff when necessary. The Florida Department of Education coordinated a state-wide search of college and university student records to help us achieve a complete postsecondary assessment for Florida study participants who had gone on to higher education. In Colorado, computer searches of postsecondary school records were conducted individually with any college or university that an interviewee reported attending. For participants from both sites, out-of-state postsecondary education records were requested when appropriate, but this response was less complete than the response for in-state records.

Records of study participants' kindergarten through 12th-grade (K–12) schooling (grades, special education, test scores, behavioral reports, suspensions, expulsions, and teachers' written comments) were collected at both study sites; these records covered 1971 through 1991. Although we collected achievement and intelligence test data (both state and national), this information was not coded, because so many different tests were found to have been used across the various districts. Study participants' GED (General Educational Development) courses and test scores (if any) and their higher education courses and grades (as mentioned earlier) were also collected. In all, we were able to obtain school records data for 89% of the study participants.

Public assistance records

To determine if a study participant had ever received public assistance, we conducted a computer search of county welfare department records at each site. Most study participants with a record of receiving assistance gave us permission to copy the pertinent information. The LTBHS study staff worked with the county welfare staff to determine the study participant's use of AFDC, food stamps, or Medicaid. Because it was not feasible for study staff to do so, public assistance records were not searched in any welfare jurisdictions outside the two study sites.

In the LTBHS study interviews (conducted mostly in 1987–89), we had inquired about the study participant's use of public assistance (if any), irrespective of where the interviewee was living at the time of the interview. Therefore, after the records search, we compared the self-reported (interview) information about public assistance with the information we had obtained from searching welfare department records (indicating public assistance usage *after* completion of all the interviews). Our analysis focused on the time frame of 1989 to 1990, when public assistance records data has been made available to us to record at each site. At both sites, only the AFDC data were consistently available and thus included in the analyses. These data were collected from September 1989 through August 1990 at the Colorado site and from January through December 1990 at the Florida site and comprised the data set that was to be utilized in the analysis.

Our comparison showed the following: Out of 622 study participants, 161 (26%), as of their interview, had reported use of public assistance; a greater number, 241 (39%) were subsequently found to have records indicating use of public assistance. There was only 1 study participant who reported use of public assistance at the time of the interview but had no county welfare department record of such use in 1990. For our analyses, we combined the interview and records data. (For more detail on the final composite variables used for the analyses, see Chapter 4.)

Criminal records

An initial computer search of the criminal records in the multi-county area of the Florida site and in the county-wide area of the Colorado site was conducted to determine whether study participants ever had a criminal record. For each individual who did have a criminal file, LTBHS study staff worked with court-records staff to record the number and type of all adult arrests, trials, convictions, and sentences (including time in prison and on probation). We also recorded the available juvenile delinquency data (such data were found for only some of the Colorado study participants; only a few Florida study participants gave us permission to record such data). However, because of the limited and inconsistent availability of this juvenile data, we did not use it in our analyses. For those study participants who were found to have ever been arrested within a study site, all of the adult criminal records data were used in the study analyses.

Criminal files indicated that 227 (36%) of the 622 study participants in Florida and Colorado had a criminal file with at least one arrest. (No out-of-state criminal records were searched, but some criminal searches conducted at the county level contained data on criminal activity of study participants in other out-of-site locations within the state, and these data we included.) Although our criminal records search for the two study sites was limited to within-state

BEST COPY AVAILABLE

records, when we took into account all of the *self-reported* information about juvenile and adult arrests and convictions (from the interview data), this added some in- and out-of-state arrests and convictions to the crime data we had obtained from our records search. This resulted in a more complete assessment of the incidence and extent of the study participants' criminal activity. (For more detail on the final composite variables used for the criminal data analyses, see Chapter 4.)

U.S. Bureau of the Census records

Because the study depended on retrospective socioeconomic data regarding the study participant's family in 1970–71, we also decided to collect, for each site, 1970 tract-level socioeconomic data from the U.S. Bureau of the Census. As explained earlier, the Florida site consisted of two urban and two rural areas. The 1970 population density in one of these urban areas was not sufficient to require tract-level data, so we employed the Census Bureau's 1980 tract-level data for this area to provide an estimate of the socioeconomic indicators. Data analyst Yu Xie collected derived tract-level data for each study site and developed several variables that he subsequently employed in his analysis of the data. The study participants' original family addresses indicated which tracts to examine. Several indicators of poverty were derived per tract—percentage living in poverty, young-adult high school dropout rate, and percentage of female-headed households. These variables provided a basis for comparing the study sites. They also provided a basis for comparing the Head Start and No–Head Start groups' original socioeconomic status.

Data Coding and Entry

Data coding was conducted at the High/Scope Educational Research Foundation under the supervision of the LTBHS study director. The coding procedures and the entry systems that we used with the interview and records data were based largely on those employed in the High/Scope Perry Preschool Project's age-27 study (Schweinhart et al., 1993). Coding manuals developed for the LTBHS study described the standard coding and computer entry procedures to be employed. This documentation of the data codes and the coding process was prepared to facilitate the analyses by the independent scholars and also any potential secondary data analyses, as described in Chapter 4. As is often the case in studies, some variables required developing specially designed coding systems (such as creating a common K–12 grading system that would be applicable across the study participants' various school systems). The occupational coding system, the most difficult and complex, is described here.

Occupational data sources

The occupational data for the study participants and their parents were derived from the LTBHS study follow-up interview. To supplement this occupational data concerning parents, information was also derived from additional, external data sources—from the study participants' original school-entry registration forms (from elementary schools) and from the initial-contact forms that were completed by the local project coordinators, locators, or interviewers when they contacted the parents or any other family members.

Data obtained from these external sources were compared with the interview data based on the study participants' recollections of their parents' occupations around 1970–71. The external-source data allowed for more-specific occupational coding, verified the study participants' recollections, and provided an occupational code when a study participant had no recollection of a parent's occupation. In cases of discrepancies regarding specific occupations, we selected the occupation that was identified by two of the three sources. When there was no agreement, we coded the occupation the parent had written on the study participant's school-entry form. If none of the three sources indicated the specific occupation but two confirmed that the parent was employed, we coded the parent as simply having been employed at the time. Although each of these data sources had a large degree of missing specific occupational data, the composite coding provided a fairly complete picture of the occupational levels of the study participants' parents and therefore the economic status of the families around 1970–71. The expectation was that particularly for the Head Start group, although parents may have been employed, the household incomes from employment would have been at or below the 1970 U.S. Bureau of the Census poverty threshold.

Occupational codes

Occupations of both parents and study participants were first coded using the U.S. Bureau of the Census codes for 1970 (see U.S. Bureau of the Census, 1971); over 400 three-digit codes for occupational titles were employed. To improve on some of the job rankings from the 1970 system, we made some adjustments to our initial coding. For this, we used the U.S. Bureau of the Census 1980 Standard Occupational Classification (SOC) system, which appears to have addressed concerns that have been raised about some of the job rankings in the 1970 system, although determining the status of jobs remains controversial (Campbell & Parker, 1983; Osborn, 1987; Stevens & Featherman, 1981). Based on the more recent coding information, some codes for the parents' occupations were moved to higher or lower classification levels to better reflect the educational requirements or income status of their jobs. For the study participants' jobs, the more recent coding system supplied us with codes for jobs that were not found in the 1970 codes; we were also able to move some codes to higher or lower classification levels to reflect typical, current educational requirements or income status. This adjusted coding of the occupational data was next reviewed by data analyst Yu Xie, who made some further suggestions, which we used in a final coding. Subsequently, Xie recoded the data into Duncan's Socioeconomic Index (SEI; O. D. Duncan, 1961) codes for the major data analyses.

THE DATA ANALYSIS APPROACH

The overall objective of the LTBHS study's analyses was to determine whether there are effects of the study participants' having attended Head Start that are evident in the K–12 school years and beyond in the form of persistence and achievement in education, work experience, and growth toward economic independence. In addition, the analyses aimed to determine the extent to which the cited early effects on High/Scope Head Start children's intellectual performance—documented in the earlier, Head Start Planned Variation study—were evident over time. However, since the HSPV study participants had not been randomly assigned to that study's model and regular Head Start groups, as discussed extensively earlier in this chapter, our data analyses also had to assess the extent of bias due to potential background differences between the LTBHS study groups. Many sources of bias, such as regional and family background differences, likely remained uncontrolled or hidden. Estimates of these kinds of bias were examined and taken into account in the major LTBHS study data analyses.

In the following, we describe the LTBHS study's analysis approach, which included employing an advisory panel along with two independent scholars, who served as the two major data analysts.

The Advisory Panel and the Independent Analysts

A technical advisory panel of experts in social science methodology and statistics guided the data analysis phase of the LTBHS study. The advisory panel members (listed on p. xv) were selected by the foundation that provided the funding for the study's data analysis. The panel provided consultation for determining the final study sample, deriving the core set of variables, selecting the data analysis models, and conducting the analyses.

As mentioned earlier, two scholars were recruited to independently analyze and interpret the LTBHS study data: Sue Marcus of the

University of Pennsylvania, a statistician with expertise in biostatistics and education, and Yu Xie of the University of Michigan, a social scientist with expertise in sociology, population studies, educational achievement, and statistics. These scholar/analysts (hereafter referred to as the study's data analysts) had no earlier association with Head Start research, the High/Scope Educational Research Foundation, or any of High/Scope's other research studies. The advisory panel, the High/Scope researchers, and the data analysts convened for three day-long meeting sessions to plan the study analyses and interpret the findings after the analyses had been completed.

The Analytic Approaches

The LTBHS study's advisory panel recommended several major analytic approaches that ranged from conventional to "cutting-edge" (e.g., selection modeling, propensity score matching). The data analysts worked with the High/Scope researchers on the formation of the final variables (see Chapter 4), then independently conducted the statistical analyses of the data, and finally presented their results to the advisory panel for review. (For the results, see Chapter 5 and the data analysts' reports in Appendixes D, E, and F.) The following considerations served as a common ground for Xie and Marcus in conducting their analyses:

Variables and predictions from the High/Scope Perry Preschool study

As an empirical basis for selecting key variables for analyses and for making predictions in the LTBHS study, the data analysts used the pattern of findings obtained in the High/Scope Perry Preschool Project's follow-up studies through age 19 and age 27 (Schweinhart et al., 1993). As discussed previously, the major advantage of the High/Scope Perry Preschool Project, which took place in one urban area with a high incidence of poverty, was that the study participants were randomly assigned to either a preschool group or a no-preschool group. On original family background and other potential sources of bias, there was only one significant difference between study groups. For these reasons, the High/Scope Perry Preschool Project's data analyses provided a good estimate of what the lives of the former preschool subjects would have been like had they not gone to preschool; the age-19 and age-27 studies' findings thus provide a valid basis for making predictions for the LTBHS study's Head Start versus No–Head Start and Model versus Regular Head Start groups. Briefly, the High/Scope Perry Preschool study found, in its age-27 follow-up, that children from low-income families who attended the preschool program, compared with a control group of children who did not attend preschool, achieved higher levels of education (especially among females), had fewer arrests, used less public assistance, and had more economic independence.

A common core of key variables

As recommended by the LTBHS study's advisory panel, the two analysts worked with a common core of covariate and outcome variables (presented in Chapter 4). Xie employed these variables in several analyses, including selection modeling, logit regression modeling, and linear regression modeling using ordinary least squares. He also supplemented the family-background data with U.S. Bureau of the Census tract-level socioeconomic data that he derived for the original neighborhoods where the study participants had lived with their families. The Marcus data analyses employed the same covariate and outcome variables in a number of analytic approaches—analysis of covariance, propensity score matching, subclassification, hidden bias analyses, and odds ratio analyses.

Statistical adjustments for original group differences

To estimate the long-term influence of Head Start by comparing outcomes of former Head Start and No–Head Start young adults, it was necessary to first estimate the influence of other factors, referred to as covariates—such as subjects' individual characteristics, family background, regional differences. Toward this end, each analyst started with an assumption originally presented by Campbell and Boruch (1975)—that Head Start was not likely to have had a negative influence. The statistical approaches that each analyst took to adjust the data for sources of bias served to increase the validity of the estimate of Head Start's long-term influence on the outcome variables.

SUMMARY OF THE STUDY METHODOLOGY

The LTBHS design has its foundation in the HSPV study of 1970–71, which conducted a comparison of study participants who attended model versus regular Head Start programs at that time. Seventeen years later, so the LTBHS study could examine the long-term effects of participation versus nonparticipation in Head Start, we used the original study groups of the HSPV study and also added groups of study participants who had no past Head Start experience. The LTBHS study, conducted at the HSPV study sites in northwestern Florida and northern Colorado, included African Americans, Hispanic Americans, and Whites/others from both urban and rural areas. This chapter has described the many challenges that were encountered and met in identifying, locating, and interviewing the LTBHS study participants and in collecting their records data from schools, welfare departments, and criminal courts.

Because the LTBHS study's Head Start Groups (those participants retained from the HSPV study) had not been randomly assigned and because the new LTBHS study participants (mainly those forming the No-Head Start groups) were enlisted as study participants when they were young adults, the LTBHS study design is quasi-experimental and includes both current and retrospective data. Consequently, there is considerable potential for bias in why and how study participants were selected and in how data were collected. However, as described in this chapter, although the design of the study presented such limitations, the data analyses, under the direction of the study investigators, the study's advisory panel, and the independent data analysts, would seek to estimate the effects of these limitations. Chapter 4 presents a description and discussion of the common core variables that the data analysts employed to estimate the long-term benefits of Head Start.

4

THE CORE VARIABLES OF THE LTBHS STUDY

To investigate the major research questions of the Long-Term Benefits of Head Start study, High/Scope researchers, in consultation with the study's two data analysts, derived a set of core variables from the study's database. This core set of variables, as recommended by the LTBHS advisory panel, consisted of a final, common set of covariates and outcome variables to be employed by the two data analysts in applying their respective analytical approaches, which they would do independently of High/Scope and of each other. The study's advisory panel also recommended limiting the outcome variables to those that most directly addressed the research questions, thereby also limiting the probability of finding significant differences due to chance.

Regarding each variable selected and derived from the LTBHS study database, this chapter provides (1) the rationale for the selection, (2) a description of the data sources, (3) a brief description of the variable, and (4) predicted findings. Appendix B summarizes how the core variables were derived and presents their final specifications.

BASIS FOR SELECTING THE CORE VARIABLES

Chapter 3 explained how we identified the data to be collected for the LTBHS study based on the types of data that had been collected for another study of young-adult outcomes—the High/Scope Perry Preschool Project, which followed up a preschool and a no-preschool group through age 19 and eventually through age 27 (Beruetta-Clement et al., 1984; Schweinhart et al., 1993). In fact, the LTBHS study, in following up its young-adult subjects, used virtually the same interview instrument and collected records data highly similar to the records data collected by both the High/Scope Perry Preschool study and the subsequent High/Scope Preschool Curriculum Comparison study, which followed up age-23 subjects who as children had experienced three different preschool curriculum models, including the High/Scope model (Schweinhart & Weikart, 1997). Therefore it made sense that the High/Scope Perry Preschool study's findings should also provide our major guidelines for selecting the core variables of the LTBHS study. The rationale for using this study's long-term findings to guide us is this: Because of the High/Scope Perry Preschool study's random assignment of subjects to its preschool and no-preschool groups, the significant differences that study obtained are likely to have provided an especially good estimate of the potential long-term effects of high-quality preschool on children from low-income backgrounds. Thus, we hypothesized that we might expect to find a somewhat similar pattern of young-adult outcomes with respect to similar variables in the LTBHS study.

Another consideration in our selection of the LTBHS study core variables was this: The LTBHS study, as explained previously, has roots in the Head Start Planned Variation study, which compared Model Head Start with Regular Head Start in the 1970s at the Florida and Colorado sites. Furthermore, High/Scope staff provided training to the teachers in the Model Head Start preschools in the Florida and Colorado sites. Thus, the High/Scope Curriculum Comparison study, which compared the High/Scope preschool curriculum with other approaches—on variables nearly identical to those used in the High/Scope Perry Preschool and the LTBHS studies—may provide especially relevant comparisons to the LTBHS study's Model Head Start versus Regular Head Start analyses.

Overall, the LTBHS study's core variables closely match the High/Scope Perry Preschool study's variables, on which researchers found significant preschool versus no-preschool differences when subjects were young adults. As described in greater detail in Chapter 3, for data on these core variables, LTBHS researchers made use of the following: (1) data from an extensive follow-up interview conducted with each study participant at approximately age 22; (2) data from an assessment of adult functional literacy, also collected at the time of the interview session; (3) data from K–12 and postsecondary educational records; (4) data from public assistance records, as applicable; and (5) data from criminal records, as applicable. Although the LTBHS study and the High/Scope Perry Preschool study differ somewhat in their respective databases, both studies derived variables from the same types of data sources and constructed variables that would yield the most complete information.

DERIVING THE CORE VARIABLES FROM THE DATA SOURCES

In consultation with the LTBHS study's two data analysts, several different types of variable preparation took place prior to the data analyses: (1) Some variables were derived from only one data source (e.g., the variable of *gender*) and therefore did not need recoding, although several data sources might have been examined to confirm a subject's status on a number of such variables. (2) Some variables were composite variables derived from several variables within one data source (e.g., from several variables pertaining to educational level) or from variables within two or more data sources (i.e., from the interview and various records). (3) Some variables were constructed by recoding to more appropriately represent the obtained frequency distribution (i.e., levels originally coded, when found to have low frequencies, were sometimes combined) or to accommodate the requirements of the various analytic approaches. As an example of the latter type of variable construction, several variables described in this chapter were set up as continuous or categorical (binary or multilevel) in preparation for one or more of a diversity of analytical models that would be employed (see McCall & Appelbaum, 1991). (4) For specific analyses, the data analysts sometimes found it necessary to combine several levels of a given variable or to further combine variables, as specified by the two data analysts Appendixes D, E, and F.

The core variables selected for the LTBHS analysis are presented in Table 4.1. These core variables consist of the *covariates* (which may mediate the outcomes of the study) and the *outcome variables* (associated with participating versus not participating in Head Start).

The remainder of this chapter is a detailed discussion of the core variables. First the covariates and then the outcome variables are explained, one at a time. Appendix B summarizes this discussion in chart form.

THE CORE COVARIATES

To examine the potential influence of the LTBHS study participants' group status, individual char-

Table 4.1: Core Variables Selected for Analyses

Type of Variable	Category for Study Participant	Variable
Covariate	Group status	• Head Start vs. No-Head Start (for FL and CO) • Model Head Start vs. Regular Head Start (for FL only)
	Individual characteristics	• Gender • Racial/ethnic membership • Age at interview • Kindergarten participation
	Family background	• Number of siblings • Household size/structure • Mother's education • Father's education • Mother's employment • Father's employment • Family welfare assistance • Site tract-level characteristics
Outcome variable	Education	• Educational attainment • Grade point average (GPA) • Years in educable mental impairment (EMI) classes • Adult APL Survey score
	Life situation/economic status	• Employment/school-enrollment status • Teen-parent status • Public assistance
	Criminal activity	• Arrests/convictions

acteristics, and family background on their later life outcomes (educational, economic, etc.), the following covariates were identified and prepared for the data analyses:

Study Participant's Group Status

Chapter 3 described the process of assigning LTBHS study participants to the study groups. To examine the question of Head Start's long-term effects, two major study groups were formed according to whether or not each study participant had attended a Head Start program in 1970–71. Also, at the Florida site, whether or not a study participant had been in a Model Head Start program (High/Scope) or Regular Head Start program was also established. (This latter type of group status goes back to study participants' having been subjects of the Head Start Planned Variation [HSPV] study.)

Head Start group versus No–Head Start group

To examine Head Start's potential contribution to later life situations or outcomes, each study participant in each site was coded for membership in the ***Head Start*** group or the ***No–Head Start*** group.

Model Head Start group versus Regular Head Start group

As described in Chapter 3, verification of participation in the HSPV study's model Head Start program could be determined only for the former Head Start children at the Florida site. Thus only members of the Florida Head Start group were able to be coded for membership in the ***Model Head Start*** group or the ***Regular Head Start*** group.

Study Participant's Individual Characteristics

Gender

For this variable, we simply coded the study participant as ***male*** or ***female***, and this coding was employed in all analyses.

Racial/ethnic membership

A racial/ethnic membership variable was constructed to describe the study participant's racial or ethnic background and to examine the potential association with other variables. Descriptive data analyses determined that there were three major groups: African Americans (at the Florida site); Hispanic Americans (at the Colorado site); and Whites/others (at both sites). This last group also included a very small number of study participants who, in the course of the interview, identified themselves as Native American, Asian American, or of other background; these latter types of study participants were included among the Whites/others, since their numbers were too few to justify creating a separate classification for each. For the analyses, the variable was thus set up to be ***African American*** or ***White/other*** in Florida and to be ***Hispanic American*** or ***White/other*** in Colorado.

Age at interview

Since the LTBHS study's interview data collection took place over approximately a 3-year period across sites, it was important to determine if the study groups differed in age at the time they were interviewed. The ***age at interview*** variable was constructed by calculating the number of months between the subject's birth date and the time of interview.

Kindergarten participation

In the early 1970s, at the study sites, kindergarten was not as available as it is today; therefore, some study participants did not attend kindergarten. Because some study participants were uncertain about whether they had attended Head Start or kindergarten or both (the terms preschool and kindergarten were used interchangeably in some places at that time), kindergarten attendance was determined by verification from school records. For the analyses, a study participant was coded as either ***no, never attended kindergarten; yes, attended kindergarten;*** or ***kindergarten attendance not known.***

Study Participant's Family Background

Most of the family-background variables used in the LTBHS study were the same as those used in the High/Scope Perry Preschool study: ***number of siblings, household size/structure, mother's education, father's education, mother's employment, father's employment,*** and ***family welfare assistance.*** In the LTBHS study, these variables are important for comparing the socioeconomic status of the Head Start and No–Head Start groups at the time of school entry.

The two studies, however, collected data for these variables in different ways. The High/Scope Perry Preschool study, as a *prospective* longitudinal study, collected family-background data at the time of the child's preschool year. By contrast, the LTBHS study is a *retrospective* follow-up study that collected the study participants' childhood family-background data when they were young adults, and thus this study is subject to the limitations of respondents' long-term recall. To achieve information that was as complete as possible, we derived family-background data in the LTBHS study from a series of questions administered at the time of the follow-up interview (when subjects were approximately 22 years old). Additional family background data were derived from three sources: (1) the brief family-contact interviews conducted with some parents to locate the study participants for the age-22 follow-up, (2) the study participants' original school-entry registration forms, and (3) U.S. Bureau of the Census tract-level data derived in accordance with the study participants' original-household addresses. The core family background variables selected for the Head Start data analysis are as follows.

Number of siblings

Because the number of siblings typically affects the family economic level, a sibling variable was constructed from two items in the age-22 follow-up interview: (1) questions asking the study participant to list younger and older siblings (including brothers, stepbrothers, sisters, and stepsisters) up to the time of the interview and (2) a question asking the study participant to specify all the members of the household when he/she was 4 to 5 years old, about the time of school entry (hereafter referred to as "original-household" data). The specific sibling variable was the study participant's ***total number of siblings.***

Household size/structure

The original household composition variables pertained to the size and structure of the household —the number of household members and whether the household was a one-parent or a two-parent household. These variables were derived from the original-household profile provided by the study participant in the age-22 interview. For this profile, the study participant specified all household members at the time of his/her school entry. The specific household-size variables were the ***total number of children*** (this included the study participant and any siblings, cousins, or other children in the household), the ***total number of adults,*** and ***the total number of people*** in the original household. The household-structure variable, derived from the household-members data, was the ***total number of parents*** (including stepparents) who lived in the original household about the time of the study participant's entry into school. This structure variable essentially allowed researchers to examine the potential effects of a one-parent versus a two-parent household.

Mother's and father's education

Derived from the study participant's interview data, the variables on mother's and father's educational level were the highest year of education completed by each. The specific variables for data analysis were ***mother's highest year of education completed*** and ***father's highest year of education completed.*** For some data analysis purposes, a second pair of categorical variables on ***mother's*** and ***father's educational level*** were formed by converting the years of education for each of them into these four levels, or categories: (1) ***unknown,*** (2) ***less than high school,*** (3) ***high school graduate/GED*** (General Educational Development Certification), or (4) ***some postsecondary education.***

Mother's and father's employment

The mother's and father's employment variables were derived from three sources: (1) In the age-22 interview, the interviewer completed a chart based on information the study participant provided about the type and duration of each parental job. (2) In the initial-contact interview that was conducted to locate the study participant for the age-22 interview, parents were asked about their type of job at the time of the child's school entry. (3) From the study participant's initial school-entry registration form, LTBHS staff obtained data indicating the parent's type of job.

To determine each parent's employment status when the study participant was approximately 4 to 6 years old, information from all three sources was compared, as explained in Chapter 3. These three data sources yielded conflicting data in only a few cases, and any conflict was resolved by giving priority to the records data, except when records data were unavailable or vague (e.g., "she worked at the telephone company") and the study participant's recall of the job was more specific (e.g., "she was a secretary at the telephone company"). The use of the three data sources reduced the amount of missing parent-employment data and also reduced our need to rely solely on the study participant's recall.

As also explained in Chapter 3, employment data were coded according to 1970 U.S. occupational classification coding system (U.S. Bureau of the Census, 1971) using over 400 (three-digit) codes for occupational titles. As explained in Chapter 3, a few jobs were recoded with reference to the U.S. Bureau of the Census 1980 Standard Occupational Classification System (SOC). The coding was then reviewed by data analyst Xie, who subsequently recoded the data into Duncan's Socioeconomic Index (SEI) codes ranging from 1–100 (O. D. Duncan, 1961) for the major data analyses. The Duncan SEI codes are considered to provide a good estimate of the status rankings of occupations. The resulting variables were ***mother's type of job*** (if employed) and ***father's type of job*** (if employed) when the study participant was 4 to 6 years of age. From the sources described above, the following categorical variable was also derived: if ***mother (father) never worked; mother (father) worked at some point during study participant's preschool, elementary, or teen years;*** or ***unknown.*** When some type of mother's (father's) employment was indicated but the data sources were not sufficient to code the type of job, that mother or father was coded as having been employed.

Family welfare assistance

Data on family welfare assistance were derived from study participants' responses to an interview question that asked them to rate how frequently, if ever, their family had been on welfare. Based on data derived from that interview question, the family's ***frequency of welfare assistance/elementary years*** was coded as ***never, occasionally, frequently,*** or ***unknown.*** A parallel variable also derived from the interview data, the family's ***frequency of welfare assistance/middle and high school (teen) years*** was similarly coded. Finally, for some analyses, a variable was constructed from both of these variables to indicate if the family's ***use of welfare assistance overall*** (across the study participant's elementary school and teen years) was ***never, occasionally, frequently,*** or ***unknown.***

Site tract-level characteristics

As described in Chapter 3, the Florida site consisted of two urban areas, each having a bordering rural area. The Colorado site consisted of a small city that was surrounded by a rural area of farms and ranches. The tract-level variables were derived by Xie from 1970 U.S. Census Bureau data, based on the study participants' home addresses at the time of school entry (or from 1980 data, for part of one site where the 1970 population density did not require tract-level census data). Although U.S. Bureau of the Census data for the sites had been used to select the No–Head Start group, it was important to provide an independent assessment of the study groups' original potential differences and at the same time derive poverty status variables for each study site and group.

The four variables derived from the tract-level data were ***tract number*** (assigned by the U.S. Bureau of the Census and employed for descriptive purposes); ***percentage of persons living in poverty; percentage of households headed by a female; percentage of youths who dropped out of high school.***

THE CORE OUTCOME VARIABLES

As explained earlier in this chapter, the outcome variables used in the LTBHS study's data analysis were selected on the basis of the High/Scope Perry Preschool Project's key significant findings ($p < .05$, unless otherwise specified). These findings are summarized in inserts that appear throughout the remainder of this chapter, as we discuss one-by-one the outcome variables employed in the LTBHS study (see again Table 4.1 on p. 53). Where appropriate, at the end of our description of an LTBHS outcome variable, we discuss expected findings on that variable vis à vis the High/Scope Perry Preschool study findings. That is, we state whether it is reasonable to expect a pattern of findings similar to that found in the High/Scope Perry study. These predictions are usually made with reference to potential Head Start effects.

To examine the young-adult effects of Head Start versus No–Head Start, and of Model versus Regular Head Start, the following outcome variables were identified and prepared for the data analysis.

Study Participant's Education

The core educational variables described in the following are the study participant's highest educational level, the study participant's grade point average (GPA), the study participant's years (if any) in educable mental impairment (EMI) classes, and the study participant's score on the Adult APL Survey of functional literacy, completed at the time of the follow-up interview.

Educational attainment

We derived educational data by combining the study participant's historical data from the interview and the school records data (including, where applicable, postsecondary education, up to 2 years after the interview); when the interview and records data disagreed, the higher reported value was accepted as the ***highest year of education completed.*** For use in some data analyses, a second, categorical variable on ***highest educational level*** was constructed by using the same four-point scale employed in constructing the mother's and father's similar education variable: (1) ***unknown,*** (2) ***less than high school,*** (3) ***high school graduate/GED,*** or (4) ***some postsecondary education.*** (Adult-education high school courses were *not* treated as postsecondary education; if the study participant had, as an adult, completed a high school diploma or the GED, the study participant's education was coded as "high school graduate/GED.") Finally, a categorical variable on ***postsecondary/college education*** was created with these two levels: (1) ***no college*** or (2) ***some postsecondary education.*** This variable was based on a study participant having completed at least some postsecondary education courses (including college courses and training for specialist certification—mechanic, computer technician, etc.).

Our expectations regarding findings on these educational attainment variables were based on the age-19 and the age-27 findings of the High/Scope Perry Preschool Project, as listed in the insert on p. 58. That study's findings include a preschool program benefit with respect to the variables *highest year of education completed* and *regular high school graduation,* with a nonsignificant trend in this same direction when the variable *regular or adult high school graduation or GED certification* was taken into account. An examination of this educational benefit finding according to gender revealed that much of the overall positive educational attainment finding may have been attributable to significantly more of the High/Scope program females having completed high school and postsecondary education.

Based on the High/Scope Perry Preschool study findings, we would expect to find in the LTBHS study a similar pattern of higher attainment in education for the Head Start subjects versus the No–Head Start subjects, especially for Model Head Start participants (who used the High/Scope Curriculum) and for female Head Start participants.

Grade point average (GPA)

A standard 1–15 coding system was developed in the LTBHS study for the study participants' A+ to F course grades collected from school records. Across school systems in the study sites, the study participants' grades for each course up to the 12th grade were converted to this standard coding system. For analysis purposes, on the advice of data analyst Xie, these standard codes were subsequently converted to a 4.3–0 coding system. (See Table B.2 in Appendix B for further details.)

To assess GPAs across major time periods, the school years were categorized into these three equal time periods: elementary school (grades 1 through 4), middle school (grades 5 through 8), and high school (grades 9 through 12). Kindergarten was not included, since at this level, the grades given, if any, were usually "satisfactory" or "unsatisfactory." Across the grade levels within each time period, an average grade was then calculated for each study participant. For the LTBHS study's data analyses, the following outcome variables were thus derived: ***mean grade point average in elementary school (GPA1); mean grade point average in middle school (GPA2);***

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: Educational Attainment and GPA

- Compared with the no-preschool group, at age 27 the preschool group showed more ***years of education completed*** (Mean = 11.9 yr vs. Mean = 11.0 yr).
- Compared with the no-preschool group, the preschool group showed, at age 27, a higher incidence of ***regular high school graduation*** (completing high school without dropping out; 66% vs. 45%) and, at age 27, a nearly significant higher incidence of ***regular or adult high school graduation or GED certification*** (71% vs. 54%, $p = .055$).
- Compared with the no-preschool females, at age 27 ***preschool females*** showed more ***years of education completed*** (Mean = 12.2 yr vs. Mean = 10.5 yr) and a higher incidence of ***regular or adult high school graduation or GED completion*** (84% vs. 35%).
- Compared with the no-preschool group, the preschool group had a higher ***grade point average in high school*** (2.1 vs. 1.7, on a four-point scale).

Note. From Schweinhart et al., 1993, pp. 58, 64, 68. Significance level for findings presented is $p < .05$ unless otherwise specified. "Preschool group" refers to study participants who had experienced the High/Scope preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

and ***mean grade point average in high school (GPA3).*** This would enable us to determine if there had been a Head Start effect in the earlier school years as well as in the high school years.

Our expectation regarding LTBHS study findings on high school grade point average *(GPA3)* was based on the High/Scope Perry Preschool study finding that high school GPA was significantly higher for the preschool than for the no-preschool group (see again the insert above). We therefore expected a similar finding in the LTBHS study for the Head Start group at each site, especially for the Model Head Start group in Florida, which had experienced the High/Scope curriculum. Furthermore, as described in Chapter 3, the Head Start Planned Variation study finding of IQ gains for participants in the High/Scope model Head Start (Smith, 1973) suggested a possibility that we would find enhanced school performance in the LTBHS study for the Model Head Start group versus the Regular Head Start group.

Years in educable mental impairment (EMI) classes

The special education outcome variable selected for the LTBHS study was ***total years in special education classes for children with educable mental impairment*** (these are classes for children assessed as having lower-than-average intellectual functioning, ***not*** for those with emotional impairment, learning disability, or speech/language impairment). In the LTBHS study, this variable was derived from school records data by totaling the number of school years (if any) that a study participant spent in special classes for children with educable mental impairment, from grade 1 through potentially grade 12. For some analyses, the data analysts constructed a categorical variable from these data to indicate if the study participant had ever or never been in an EMI classroom.

Again, our expectation regarding this core variable was based on special education findings in the High/Scope Perry Preschool follow-up of young adults. As is shown in the insert below, that study found that students who had attended the preschool program, compared with students having no preschool, spent significantly fewer *school years in EMI programs;* also, fewer children in the preschool group were *ever placed in EMI programs*—and this was especially true for the preschool females.

There was also, however, the High/Scope Perry study finding that the children who had attended the High/Scope preschool program tended to spend more *years in compensatory education programs* (remedial programs) than their no-preschool counterparts did. It may be that teachers or specialists referring the children to such programs perceived greater potential for the preschool participants (versus the no-preschool participants) to benefit from such services, or perhaps the preschool participants were more responsive to the screening assessment for such services. Because it is difficult for us to interpret this compensatory education finding of the High/Scope Perry study, the decision was made, in the LTBHS study, to focus on only a single special-education outcome variable, ***total years in special education classes for children with EMI.*** Although the compensatory education finding made it difficult for us to make any prediction of findings on general special education, on the EMI variable, we did expect to have findings similar to the High/Scope Perry findings, for Head Start versus No–Head Start subjects.

Adult APL Survey

To assess the study participant's literacy skills, a core outcome variable was derived from the study participant's responses to the 40 literacy items on the Adult APL Survey (American College Testing Program, 1976). The APL Survey assesses competence with facts and terms, reading, writing, computation, and problem solving. The core variable was the ***APL total score,*** which consisted of the total of correctly answered items compiled across the study participant's scores on the subtests (i.e., community resources, occupational knowledge, consumer economics, health information, government and law, facts and terms, reading, writing, computation, problem solving).

Our expectations for this core outcome variable were based on the findings from the

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: Educable Mental Impairment (EMI) and Remedial Education Services

- Compared with the no-preschool group, the preschool group spent fewer ***school years in EMI programs*** (Mean = 1.1 yr vs. Mean = 2.8 yr); fewer members of the preschool group spent ***6 or more years in EMI programs*** (9% vs. 27%), and fewer were ***ever placed in an EMI program*** (15% vs. 34%).
- Compared with no-preschool females, fewer ***preschool females*** were ***ever enrolled in EMI programs*** (8% vs. 37%).
- Compared with the no-preschool group, the preschool group spent more ***years in compensatory education programs*** (i.e., remedial services in reading, math, and so on; Mean = 1.0 yr vs. Mean = 0.4 yr).

Note. From Schweinhart et al., 1993, pp. 61–62, 64–65. Significance level for findings presented is $p < .05$. "Preschool group" refers to study participatns who had experienced the High/Scope Preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

High/Scope Perry Preschool study, in the insert below. At age 19, when compared with the no-preschool group, the preschool group had significantly higher *APL total scores* as well as significantly higher scores on a few of the APL subtests. By contrast, at age 27, the APL Survey scores were higher for the preschool group only on two of the subtests. The preschool group's mean total score was virtually the same at age 27 as at age 19, whereas the no-preschool group's mean total score had increased between those ages. Given the High/Scope Perry Preschool study's mixed findings across APL subtest scores, for the LTBHS study, only ***APL total score*** was selected as a core variable. On this variable, we expected the LTBHS study to find an advantage for the Head Start groups versus the No–Head Start groups, and especially for the Model Head Start group versus the Regular Head Start group at the Florida site.

Life Situation/Economic Status

In many ways, the LTBHS study participants' life situations at the time of the follow-up interview are good indicators of where these young people stand in the transition from youth to adulthood. Whether or not an individual is employed, in school, married, or a parent not only describes current status but also may predict the future. This was borne out in the High/Scope Perry Preschool study, which found a substantial correlation between the age-19 and age-27 variables, especially for economic and educational status. The following life-situation outcome variables were selected for the LTBHS study's data analyses.

Employment/school-enrollment status

The High/Scope Perry Preschool follow-up of young adults at age 19 looked at the study participant's employment/unemployment since leaving high school. But because the LTBHS study participants were interviewed at varying age points after their high school years, determining months of employment since high school was problematic. Thus, for the LTBHS study analyses, we decided to utilize only true employment status *at some point during the interview year,* combined with consideration of whether or not the person was enrolled in school during that year. This would describe the study participant's current life

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: The APL Survey

- Compared with the no-preschool group, the preschool group at age 19 had significantly higher scores on ***general literacy,*** i.e. ***APL total scores*** (out of 40 points—Mean = 24.6 points vs. Mean = 21.8 points). The preschool group also had significantly higher scores on the subtests of ***occupational knowledge, health information,*** and ***reading,*** and they had nearly significant higher scores on the ***writing*** subtest ($p = .056$).
- Compared with the no-preschool group, the preschool group at age 27 had significantly higher scores than the no-preschool group on the subtests of ***health information*** and ***problem solving*** but no significant differences on other subtest scores or on APL total scores—Mean = 24.8 points for the preschool group vs. Mean = 23.1 points for the no-preschool group.

Note. From Schweinhart et al., 1993, p. 67. Significance level for findings presented is $p < .05$ unless otherwise specified. "Preschool group" refers to study participants who had experienced the High/Scope preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

situation with respect to economic independence or potential for economic independence (as evidenced by enrollment in education). We therefore created a composite variable from two variables derived from the interview database.

First, whether the study participant was ***employed*** or ***not employed*** at the time of the interview or at some point during the interview year was derived from the interview responses to a series of questions about employment history. Next, whether the study participant was ***enrolled in school*** or ***not enrolled in school*** (meaning in an educational program such as GED preparation, a postsecondary program, or college) at the time of the follow-up interview or at some point during the interview year was derived from the interview responses to a series of questions about educational history since high school.

The composite variable ***employment/school-enrollment status*** was then created to refer to the study participant's in-school status (whether enrolled in school at some point during the interview year) and employment status (whether employed at some point during the interview year). For use in various potential analyses, ***employment/school enrollment status*** was also determined for each year from age 18 up to the time of the interview. The final core composite variable was coded as (1) ***employed, but not in school,*** (2) ***not employed, but enrolled in school,*** (3) ***employed and in school,*** or (4) ***not employed and not in school.*** For use in some analyses, a categorical variable of employment/school-enrollment status was constructed to indicate if the study participant was ***not employed and not in school*** or instead was ***in school or employed or both*** (i.e., in any one of situations 1, 2, or 3).

Our expectations for the composite employment/school enrollment variable were based on the findings of the High/Scope Perry Preschool study, as shown in the insert below. In that study, significantly more members of the preschool group than of the no-preschool group were *employed* at age 19. Also, members of the preschool group, had more *months of employment* in their 19th year. Later, when compared with the no-preschool group at age 27, the preschool group had a nearly significant higher educational level ($p < .055$), although there were no significant differences between groups in incidence of employment. This pattern of education and employment found in the High/Scope Preschool study suggested that in the LTBHS study we might expect more Head Start than No-Head Start study participants to be employed or in school at approximately age 22.

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: Employment

- Compared with the no-preschool group, a significantly greater percentage of the preschool group were **employed** at the time of the age-19 interview (50% vs. 32%).
- Compared with the no-preschool group, the preschool group reported significantly more **months of employment** in the age-19 calendar year (Mean = 6.1 mo vs. Mean = 3.9 mo).
- Compared with the no-preschool group, the preschool group reported at age 19 significantly fewer **months unemployed** since leaving school (Mean = 4.9 mo vs. Mean = 10.3 mo).

Note. From Schweinhart et al., 1993, p. 115. Significance level for findings presented is $p < .05$ unless otherwise specified. "Preschool group" refers to study participants who had experienced the High/Scope preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

Teen-parent status

The family formation issues examined in the LTBHS study pertain to whether the study participant had become a parent as a teenager. The selected variable was derived from the study participants' responses to interview questions that asked them to indicate the dates of their children's births (if any). From these data, along with each study participant's own birth date and the date of the interview, we determined how old the study participant was at the time of each child's birth (if any).

The derived core variable for the data analysis was ***teen-parent status,*** indicating whether the study participant, at less than age 20, ***had*** or ***did not have a child born*** (regardless of subsequent parenthood). A variable concerning whether the study participant ever became a parent when less than age 20 and *also not married* could likewise have been generated from the LTBHS study data, since the interview did include questions about the dates of any marriages or divorces. However, the data analysts advised that we select only the ***teen-parent status*** variable as a core outcome variable. Because there has been a large increase across socioeconomic groups in young people living with a spouse-equivalent and having children without being married, the findings on a variable pertaining to unmarried teen parenthood, although certainly relevant, would be difficult not only to predict but also to interpret, if group differences were found (U.S. Bureau of the Census, 1989).

As shown in the insert below, the High/Scope Perry Preschool study found no significant preschool program effect on the variables relevant to teen-parent status but did find fewer births to the preschool group's unmarried females up to age 27. Given these findings along with our decision to focus on teen-parent status irrespective of marital status, it was difficult to make any predictions on the core outcome variable relating to parenthood for the LTBHS study.

Public assistance

The core variable on public assistance pertained to whether the study participant had ever received any form of public assistance to meet economic needs. The data sources for the composite public assistance variable were (1) the study participant's responses regarding currently receiving public

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: Marital and Parental Status

- At age 19, no significant differences between preschool and no-preschool groups were found for males or females in ***average number of births:*** By that age, 44% of the preschool females versus 54% of no-preschool females had given birth to one or more children; 18% of preschool males versus 26% of no-preschool males had fathered one or more children.
- No significant differences were found between the preschool group and the no-preschool group at age 19 for ***marital status;*** of the overall sample, 89% were not married.
- Up to age 27, compared with no-preschool females, preschool females reported a significantly smaller percentage of their ***births*** to have occurred ***while unmarried*** (57%, or 26 out of 46 births, vs. 83%, or 43 out of 52 births).

Note. From Schweinhart et al., 1993, pp. 122–123. Significance level for findings presented is $p < .05$ unless otherwise specified. "Preschool group" refers to study participants who had experienced the High/Scope preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

assistance (AFDC, food stamps, General Welfare/Assistance, or Social Security) at the time of interview and (2) county welfare records of current AFDC recipients between September 1989 and August 1990 (after the interviewing phase of the study had ended). Because only this single year of public assistance records data was made available at the Florida site, this was treated as the target year for both sites for the variable derived from records data.

As Chapter 3 described, the interview data indicated a lower percentage of study participants receiving public assistance than was indicated by the records data collected from 1 to several years later; perhaps fewer had indeed been using public assistance at the earlier point, or perhaps some interviewees chose not to indicate their use of public assistance. Also as noted in Chapter 3, when the study interview and public assistance records data were combined, this added only 1 name to the list of study participants who had received public assistance according to the 1989–90 records data. Although the records data pertained only to AFDC usage, the higher usage of public assistance found in these data suggests that AFDC usage was highly indicative of (1) study participants' previous use of public assistance and (2) study participants' use of other types of public assistance. Therefore, the composite appears to be indicative of study participants' ever having received public assistance as young adults.

The public assistance core composite variable was coded as (0) ***never used public assistance*** or (1) ***used public assistance.*** This variable was based on evidence of the study participant's having used AFDC public assistance from September 1989 to August 1990 and the study participant's having reported use of public assistance of any kind at the time of interview.

Our expectations for findings on this public assistance variable were based on related findings of the High/Scope Perry Preschool study, which are listed in the insert below. That study found some differences in public assistance usage at age 20–24 and in the 10 years prior to age 27, with the preschool group having received less public assistance than the no-preschool group. By age 27, various public assistance sources were further investigated, and this revealed a continuation of the earlier pattern of the preschool group using less public assistance. Thus, in the LTBHS study, with study participants interviewed at 22 years of

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: Use of Public Assistance

- Compared with the no-preschool group, the preschool group had significantly fewer members (17% vs. 38%) on ***General Assistance and food stamps*** at ages 20 to 24.
- At age 27, compared with the no-preschool group, the preschool group had fewer members (59% vs. 80%) who had received social services assistance sometime in the previous 10 years; this assistance was primarily ***AFDC, food stamps,*** and ***General Assistance,*** plus in a few cases, ***protective services, Medicaid,*** and ***public housing.***
- At age 27, compared with no-preschool females, significantly fewer preschool females ***were receiving food stamps*** (21% vs. 50%).

Note. From Schweinhart et al., 1993, pp. 107, 109, 114. Significance level for findings presented is $p < .05$ unless otherwise specified. "Preschool group" refers to study participants who had experienced the High/Scope preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

age and public assistance records data collected 1 to 2 years later, we would expect to find across study groups an indication of somewhat greater public assistance usage than in the High/Scope Perry Preschool study at age 19. We also expected to see the beginning of a pattern of less public assistance usage by the Head Start groups versus the No–Head Start groups. Given the High/Scope Perry study's finding of a significant reduction in food stamp usage by preschool females, we might expect less use of public assistance by Head Start females versus No–Head Start females.

Criminal Activity—Arrests and Convictions

Another set of core outcome variables for the LTBHS study data concerned the arrests and convictions (if any) of the study participants. These variables were based on the interview data obtained at approximately age 22 and criminal records data collected in the area of each study site 1 to 2 years after the interview. In the interview, the study participant was asked whether he or she had ever been arrested for committing a crime and, if so, to list up to three of the most serious arrest incidents and up to three of the most serious convictions, indicating where these took place. (The term *arrest incidents* refers to the number of times a person was arrested, hereafter referred to as the number of arrests.) We were then able to determine whether or not each arrest or conviction had occurred within one of the counties or the local court jurisdictions from which we had collected criminal records data.

The arrests and convictions from the interview data were added to the criminal records data if any *one* of the following conditions existed:

1. The study participant's repeated arrests or convictions occurred before age 18, so that they might not have been encountered in our search of criminal records (we did not set out to include court juvenile records in our search, because they were not uniformly available across the study participants at both sites).

Findings From Age-19 and Age-27 Follow-ups of Participants in the High/Scope Perry Preschool Study: Criminal Activity

At age 19 and age 27, compared with the no-preschool group, the preschool group was found to have engaged in significantly less criminal activity as indicated by the following group averages or percentages:

- At age 19, a lower incidence of ***ever arrested*** (31% vs. 51%)
- At age 19, a lower incidence of ***ever arrested for adult misdemeanors*** (2% vs. 16%)
- At age 19, fewer ***felony arrests*** (0.7 vs. 2.0)
- A lower ***number of adult arrests*** (1.3 vs. 2.3 at age 19; 1.8 vs. 4.0 at age 27)
- Fewer ***adult misdemeanor arrests*** (0.02 vs. 0.33 at age 19; 1.2 vs. 2.5 at age 27)
- At age 27, fewer ***undropped misdemeanor cases (0.8 vs. 1.4)***
- At age 19, from the self-report misconduct scale items, fewer times ***involved with the police*** (0.6 vs. 1.0); no significant findings on this variable at age 27

Note. From Schweinhart et al., 1993, pp. 84, 86, 91, 92, 95. Significance level for findings presented is $p < .05$ unless otherwise specified. "Preschool group" refers to study participants who had experienced the High/Scope preschool program; "no-preschool group" refers to study participants who had experienced no preschool program.

2. The arrests or convictions occurred at age 18 or beyond and also inside the areas of jurisdiction where we collected criminal records data, but they were erroneously *not* included in the records.
3. The arrests or convictions occurred at age 18 or beyond, outside of the areas of jurisdiction covered by our criminal records data collection.

By compiling the records data and interview data, we created two composite variables: ***total arrests,*** meaning the study participant's total number of arrests (arrest incidents), including those for any misdemeanors or felonies, and ***total convictions,*** meaning the study participant's total number of convictions, including those for any misdemeanors or felonies. (See Table B.2 in Appendix B for further details.)

As is shown in the insert on p. 64, the High/Scope Perry Preschool study found a consistent pattern of less criminal activity for the preschool group versus the no-preschool group. Therefore, for the LTBHS study, we expected such a pattern to emerge for the Head Start subjects versus the No–Head Start subjects by the time the study participants were interviewed at approximately age 22 and by the time their criminal records data, if any, were collected 1 to 2 years later. Since the pattern of High/Scope Perry findings was so consistent across a range of crime-related variables, we expected that any differences in criminal activity between the LTBHS study groups would be sufficiently indicated by the number of arrests, convictions, or both.

SUMMARY

The LTBHS study covariates described in this chapter provide a basis for adjusting for some of the initial Head Start versus No–Head Start group differences and thus should contribute to an estimate of the long-term effects of the Head Start experience. However, as indicated here and in Chapter 3, because the No–Head Start group in the LTBHS study was not randomly selected, we anticipated that this comparison group would give a limited estimate of what the former Head Start students' lives would have been had they not attended Head Start. Furthermore, although the LTBHS study's covariates are comparable to those employed in the High/Scope Perry Preschool study, the family-background covariates of the two studies were derived differently. Unlike those of the High/Scope Perry study, the family-background covariates of the LTBHS study had to be derived retrospectively, from the follow-up interview that was conducted when the LTBHS study participants were young adults; these covariates were thus subject to the limitations of recall. Nevertheless, the expectation was that we would find, for the LTBHS Head Start groups, a somewhat similar pattern in terms of outcome variables as was found for the preschool group in the High/Scope Perry Preschool study. Also, because the Model Head Start group from the LTBHS study used the High/Scope curriculum model, we expected to find greater benefits for that group versus the Regular Head Start group. The rationale for these expectations is fully discussed in Chapter 5. Detailed analyses of the study designs constraints are also provided in the data analysts' reports in Appendixes D, E, and F.

5

FINDINGS ON THE ORIGINAL COMPARABILITY OF THE LTBHS STUDY GROUPS

The Head Start and No–Head Start study participants in the Long-Term Benefits of Head Start (LTBHS) study did not differ on the core covariates of gender, kindergarten participation, or age at interview. However, as expected, the Head Start study participants predominantly came from neighborhoods that were more impoverished economically and from households of significantly lower socioeconomic status; more than the No–Head Start group, the Head Start group tended to come from single-parent households; they also came from households with more children, and their fathers had lower educational levels and lower scores on Duncan's Socioeconomic Index of job types. These findings, on the one hand, demonstrate that Head Start programs had indeed effectively recruited their participants from the most economically impoverished families. On the other hand, these family-background differences of the LTBHS study groups presented limitations when it came to estimating Head Start's potential benefits. Although we found the Model Head Start group versus the Regular Head Start group to have few of the family-background differences mentioned above, there were, for these two subgroups at the Florida site, differences in the poverty levels in their respective geographic areas. These differences presented limitations regarding estimation of the potential benefits of the Model Head Start program (which was the child development, theory-based High/Scope program).

The Long-Term Benefits of Head Start study is the first major study that has followed former Head Start children into their young-adult years. Thus, educators, policymakers, and others will be interested in its findings about how many former Head Start children graduated from high school, went on to postsecondary education, or were on the road toward economic independence by age 22. Although such descriptive statistics by themselves are of interest, they take on even more meaning alongside an estimation of what these young adults' lives would have been like without Head Start. As described in earlier chapters, to provide such an estimate, this study compared some young adults who had participated in Head Start with those who had not. Although the No–Head Start study participants were from a socioeconomic background roughly comparable to that of the Head Start study participants, the findings presented in this chapter indicate that the No–Head Start group did come from families that were somewhat more advantaged economically.

This chapter first provides an overview of the analysis process. Second, based on the work of data analysts Sue Marcus and Yu Xie, the chapter presents the major findings from comparing the Head Start versus No–Head Start groups as well as the Model versus Regular Head Start groups on the study's core covariates—the study participants' individual characteristics and family backgrounds (including the site tract-level characteristics). The chapter concludes with a discussion of how the findings on covariates influenced the expected findings on the study's core outcome variables—the variables having to do with educational attainment and other life situations of study participants.

To supplement the information provided in this chapter, Appendix C presents summary tables of some specific findings about the core covariates. Also, reports by Xie (Appendixes D and E) and Marcus (Appendix F) provide further details of the analyses they conducted, the findings they obtained on the core variables, and their interpretations of these findings.

AN OVERVIEW OF THE ANALYSIS PROCESS

Because Head Start recruitment procedures aim at enlisting families of the lowest socioeconomic status, we realized from the outset of the study that the Head Start study participants could be expected to be more disadvantaged economically than the No–Head Start study participants. One national study of Head Start recruitment criteria and procedures (Love & Grover, 1987) found that Head Start agencies have indeed usually focused on recruiting families with the greatest economic need. Therefore, since the LTBHS study design was quasi-experimental, a major challenge for the study's data analysts would be to adjust for any findings of initial differences between the LTBHS study groups. Several methodologists, such as Campbell and Boruch (1975), have pointed out that in program evaluation research, a program's effects are often underestimated because of such biasing influences as initial differences on family-background variables. However, as pointed out by Manski and Garfinkel (1992), Xie, Marcus, and others, even experimental studies with random assignment to study groups have their limitations, and quasi-experimental studies (such as the LTBHS study) can serve to provide useful insights as well as to corroborate findings from experimental studies.

As indicated previously, to deal most effectively with the situation of the LTBHS study's nonrandom design, we consulted a panel of nationally known experts in research methodology and statistics. This panel recommended that in addition to conducting traditional statistical analyses of study data, we employ selection modeling, examine U.S. Census tract-level socioeconomic data, and conduct analyses using propensity scores. The study's independent data analysts—Sue Marcus and Yu Xie—who initially provided consultation to the High/Scope researchers on the selection and creation of core study variables (as described in Chapter 4), subsequently conducted analyses on the original comparability of the study groups and then adjusted for the obtained initial group differences they found, in order to most closely estimate the potential long-term benefits of Head Start. The following paragraphs provide a brief description of each of these steps.

Selecting the LTBHS Study's Variables and Predicting Findings

As detailed in Chapter 4, the pattern of findings obtained in the High/Scope Perry Preschool study (Schweinhart et al., 1993) served as the empirical basis for our selection of core covariates and outcome variables from the LTBHS study's database and also for our predictions about study findings. As discussed previously, the major advantage of the High/Scope Perry Preschool study (which took place in one urban area with a high incidence of poverty) was that the study participants had been randomly assigned to either the High/Scope preschool group or the no-preschool, control group. There was only one significant difference between the study groups on original family background (slightly more employment among the single mothers in the no-preschool group) and little or no other detectable potential source of bias.

Examining the LTBHS Study Groups' Initial Comparability

Before estimating the long-term influence of Head Start by comparing the Head Start and No–Head Start groups on the core outcome variables, it was necessary to estimate the influence of the factors referred to as covariates. These factors have to do with the study participant's **individual characteristics** and **family background** (including the tract-level characteristics of the geographic area of the participant's original household).

To examine the factor of tract-level characteristics, Xie generated geographic maps of the various areas within each site where the study participants' 1970–71 family households were located. He then identified the corresponding U.S. Bureau of the Census tracts and derived tract-level socioeconomic data for each study site area, to examine the nature of the potential differences in site characteristics. In a sequence of regression models, he employed the other covariates with respect to each core outcome variable. Marcus examined study participants' individual characteristics and family backgrounds, developed propensity scores based on the study's core variables, and subsequently included the propensity scores in a range of statistical analyses that adjusted for the obtained initial differences between study groups. Both analysts also examined the likelihood of unobserved influences or hidden bias in the data.

Adjusting for Initial Group Differences

Although the nonrandom design of the LTBHS study precluded finding *conclusive* evidence of Head Start's long-term benefits, both analysts nevertheless found that after adjustment for initial group differences, the data do provide at least *plausible* evidence of such benefits. As supported by other program evaluation approaches (e.g., Campbell & Boruch, 1975; Manski, 1995), in making estimates of Head Start's long-term benefits, each analyst started with the assumption that being in Head Start would not diminish participants' later life-status but would instead either help participants to overcome their initial low socioeconomic condition or would have no effect. To test this assumption, in the first-stage descriptive data analyses on the core outcome variables, the data analysts each employed statistical models that did *not* adjust for the original socioeconomic differences between the Head Start group and the No–Head Start group. Next, they demonstrated that when their data analyses adjusted for the socioeconomic differences, the initially obtained finding of an advantage for the No–Head Start group generally shifted toward a finding of no differences between the groups or to a finding of an advantage for Head Start.

THE FINDINGS ON THE COVARIATES: INDIVIDUAL CHARACTERISTICS AND FAMILY BACKGROUND

In reporting the findings on the initial comparability of the study groups, we begin with the Head Start versus No–Head Start comparability findings; these are followed by the Model Head Start versus Regular Head Start comparability findings (at the Florida site). (It should be noted that the Model Head Start program used the High/Scope preschool approach, which was also employed in the High/Scope Perry Preschool study, in the High/Scope Preschool Curriculum Comparison study, and in the 1970 Head Start Planned Variation study at the Florida and Colorado sites.)

The major sections of this chapter provide bulleted, boldface-type statements of the specific findings. These specific findings are periodically summarized along with their implications. Because of the population differences between the Florida and Colorado study sites, separate analyses were conducted for each site: Florida and Colorado findings are thus presented separately (generally, the Florida findings are presented first). Percentages and proportions in the text and in some tables have been rounded to whole numbers when appropriate. Throughout the chapter, the significance level is $p < .05$ for differences between groups unless otherwise stated. It should also be noted that in the text and tables, *M* indicates "mean" and *SD* indicates "standard deviation."

To examine the initial (1970–71) comparability of the study groups, the analysts employed the core covariates that pertained to individual characteristics of the study participants and to their family backgrounds, including the site tract-level characteristics for the area of their original households. As indicated earlier, these core covariates are based largely on interview data, with verification from school records and U.S. Bureau of the Census data. The following summarizes the Head Start versus No–Head Start and the Model Head Start versus Regular Head Start findings for each covariate.

Study Participants' Individual Characteristics

Comparing the Head Start group and the No–Head Start group at the two sites

Head Start study participants were first compared with No–Head Start study participants on *gender.*

Table 5.1: Head Start and No–Head Start Groups by Site, Gender, and Race/Ethnicity

Variable/Site	*N*	Head Start *N*(%)	No–Head Start *N*(%)
Gender			
Florida	**424**	**160**	**264**
Male	212 (50.0%)	81 (50.6%)	131 (49.6%)
Female	212 (50.0%)	79 (49.4%)	133 (50.4%)
Colorado	**198**	**130**	**68**
Male	107 (54.0%)	72 (55.4%)	35 (51.5%)
Female	91 (46.0%)	58 (44.6%)	33 (48.5%)
Racial/ethnic membership			
Florida	**424**	**160**	**264**
African American	337 (79.5%)	125 (78.1%)	212 (80.3%)
White/other	87 (20.5%)	35 (21.9%)	52 (19.7%)
Colorado	**198**	**130**	**68**
Hispanic American	117 (59.1%)	95 (73.1%)	22 (32.4%)
White/other	81 (40.9%)	35 (26.9%)	46 (67.6%)

Note. At the Florida site, the African American group included 329 participants who identified themselves as African American plus those study participants who identified themselves as Hispanic American ($n = 4$) or African American and White ($n = 4$). At the Colorado site, the Hispanic American group included those study participants who identified themselves as Hispanic American ($n = 112$) or Hispanic American and White ($n = 5$). Across the Florida and Colorado sites, the White/other group included 132 study participants who identified themselves as White plus those study participants who identified themselves as Asian American or Pacific Islander ($n = 2$), American Indian ($n = 8$), or combinations of other ethnic groups ($n = 26$).

racial/ethnic membership, kindergarten participation, and *age at interview.* Table 5.1 presents each site's percentages for *gender* and *racial/ethnic membership.* Table 5.2 summarizes the statistical comparisons between the study groups on all of the variables related to study participants' individual characteristics.

➤ **At both the Florida site and the Colorado site, Head Start and No–Head Start groups did not differ significantly on *gender,* that is, in proportions of males and females.** Table 5.1 shows the numbers and percentages of males and females in the study groups at the Florida and Colorado sites. The Florida Head Start group had 51% males and 49% females, compared with the No–Head Start group having 50% males and 50% females. The Colorado Head Start group had 55% males and 45% females, compared with the No–Head Start group having 51% males and 49% females.

➤ **Head Start and No–Head Start groups at the Florida site did not differ significantly in proportions of different *racial/ethnic membership,* but at the Colorado site, they did differ significantly on this covariate.** Table 5.1 shows that in Florida the Head Start and No–Head Start groups had roughly comparable proportions of African American study participants and White/other study participants. However, in Colorado, the study groups were significantly different in their proportions of members who were Hispanic American versus White/other. Of the Head Start study participants, 73% were Hispanic American, and 27% were White/other. In the No–Head Start group, we find the opposite pattern—32% Hispanic American and 68% White/other. In this part of Colorado in 1970–71, the lowest-income community involved was largely Hispanic American, and the White/other study participants were thus a small proportion of the children recruited for Head Start. By contrast, the remaining pool available for selecting the comparison (No–Head Start) group was largely White/other. We expected this imbalance to be an advantage for the No–Head Start group because, compared with their Hispanic counterparts, study participants who were White/other could possibly be expected to have faced less discrimination as students and to have had greater opportunities for employment after their high school years.

➤ **At both the Florida site and the Colorado site, Head Start and No–Head Start groups did not differ significantly in proportions of study participants with *kindergarten participation.*** Of Florida's Head Start study

Table 5.2: Head Start Versus No–Head Start Groups on Study Participant Characteristics

Study Participant's Characteristics	Florida	Colorado
Gender	HS = NHS	HS = NHS
Racial/ethnic membership	HS = NHS	HS > NHS
Age at interview	HS = NHS	HS = NHS
Kindergarten participation	HS = NHS	HS = NHS

Note. HS = Head Start; NHS = No–Head Start. HS = NHS indicates no significant group difference. Group differences were tested using *t*-tests and χ^2 tests, as appropriate. Greater than (>) and less than (<) indicate obtained significant differences at $p < .05$. In Colorado, "HS > NHS" means that, in Colorado, there were more Hispanic Americans in the Head Start group than in the No–Head Start group, $p < .001$.

participants, 61% attended kindergarten, and 31% did not attend kindergarten; for 8%, kindergarten participation could not be confirmed. Of Florida's No–Head Start study participants, 63% attended kindergarten, and 35% did not attend kindergarten; for 2%, kindergarten participation could not be confirmed. At the Colorado site, 87% of the Head Start group attended kindergarten and 13% were unconfirmed on this variable; 85% of the No–Head Start group attended kindergarten and 15% were unconfirmed.

➤ **At both the Florida site and the Colorado site, Head Start and No–Head Start groups did not differ significantly on *age at interview.*** The study participants were approximately 5 to 6 years old when they entered school in 1971. When the final Head Start and No–Head Start study samples were relocated and interviewed, their ages at the time of the follow-up interview were compared. In both Colorado and Florida, the average age for Head Start and No–Head Start interviewees alike was about 22 years. In Florida, the average age for the Head Start group, was 22.2 years (SD = 1.5); for the No–Head Start group, it was 22.2 years (SD = 1.3). In Colorado, the average age for the Head Start group was 22.1 years (SD = 1.0); for the No–Head Start group, it was 22.1 years (SD = 1.1).

Summing up individual characteristics—Head Start versus No–Head Start

At both the Florida site and the Colorado site, there were no Head Start versus No–Head Start differences in study participants' individual characteristics of gender, kindergarten participation, *or* age at interview. *In Florida, there were also no differences between the two groups on* racial/ethnic membership. *However, in Colorado, there was a higher proportion of Hispanic Americans in the Head Start group than in the No–Head Start group. Thus, interpretations of the Colorado findings would have to take into account how this racial/ethnic difference might affect generalizations.*

Comparing the Model Head Start group and the Regular Head Start group at the Florida site

➤ **At the Florida site, the Model Head Start group did not differ significantly from the Regular Head Start group on either *gender* or *racial/ethnic membership.*** For gender, the Model Head Start group (n = 74) had 45% males and 55% females, compared with the Regular Head Start group (n = 86) having 56% males and 44% females. For racial/ethnic membership, the Model Head Start group was 73% African American and 27% White/other, compared with the Regular Head Start group being 83% African American and 17% White/other.

➤ **Florida's Model Head Start and Regular Head Start groups differed significantly in proportions of study participants with *kindergarten participation.*** In Florida, 97% of the Model Head Start group attended kindergarten, with 3% unconfirmed on this variable. In comparison, 30% of Florida's Regular Head Start group attended kindergarten, and 58% did not attend kindergarten, with 12% unconfirmed.

➤ **Florida's Model Head Start and Regular Head Start groups did not differ significantly on *age at interview.*** The study participants were all approximately 5 to 6 years old when they entered school in 1971. When they were relocated and interviewed as young adults, the age at interview averaged 22.0 years (SD = 1.7) for the Model Head Start group and 22.3 years (SD = 1.3) for the Regular Head Start group.

Summing up individual characteristics—Model Head Start versus Regular Head Start

Florida's Model Head Start group (using the High/Scope preschool curriculum) and Regular Head Start group did not differ on gender, racial/ethnic membership, *or* age at interview. *They did differ, however, on* kindergarten participation: *The Model Head Start group had the likely advantage of most members having a year of* kindergarten participation, *whereas approximately only one third of the Regular Head Start group had a year of kinder-*

garten. This difference would have to be taken into account in interpreting data analysis findings on the core outcome variables for the Model Head Start group versus the Regular Head Start group.

Study Participants' Family Backgrounds

Comparing the Head Start group and the No–Head Start group at the two sites

The Head Start group was next compared with the No–Head Start group on these family-background variables: total *number of siblings* (over lifetime), original *household size/structure, mother's education, father's education, mother's employment, father's employment, family welfare assistance.* The findings from analyses on these family background variables, reported by both Xie and Marcus for Head Start versus No–Head Start comparisons, are summarized in Table 5.3. (Appendix C presents additional tables on some of these variables.)

➤ **At each of the two sites (Florida and Colorado), Head Start families had a significantly greater *number of siblings* than did No–Head Start families.** At the Florida site, the number of siblings (over the course of a study participant's life) was significantly greater for the Head Start group than for the No–Head Start group (M = 4.2 [SD = 2.7] vs. M = 3.6 [SD = 2.6]). At the Colorado site, the Head Start group also averaged a significantly greater number of siblings than did the No–Head Start group (M = 3.9 [SD = 2.5] vs. M = 2.6 [SD = 1.8]).

➤ **At each of the two sites (Florida and Colorado), compared with the original households of the No–Head Start group, the original**

Table 5.3: Head Start Versus No–Head Start Groups on Family Background

Study Participant's Family Background	Florida	Colorado
Number of Siblings	HS > NHS	HS > NHS
Household size/structure:		
Number of children	HS > NHS	HS > NHS
Number of adults	HS < NHS	HS = NHS
Mother's education:		
Completed high school/GED	HS = NHS	HS < NHS
Some postsecondary education	HS = NHS	HS < NHS
Highest year of education completed	HS = NHS	HS < NHS
Father's education:		
Completed high school/GED	HS < NHS	HS < NHS
Some postsecondary education	HS = NHS	HS < NHS
Highest year of education completed	HS < NHS	HS < NHS
Mother's type of employment	HS = NHS	HS<NHS*
Father's type of employment	HS = NHS	HS< NHS
Family welfare assistance (elem. school)		
Never on welfare	HS < NHS	HS < NHS
Frequency on welfare	HS > NHS*	HS = NHS

Note. HS = Head Start; NHS = No–Head Start. HS = NHS indicates no significant group difference. Mother's and father's type of employment were classified according to Duncan's Socioeconomic Index (SEI) scores. Group differences were tested using t-tests and χ^2 tests, as appropriate. Greater than (>) and less than (<) indicate obtained significance at $p < .05$.
*$p < .10$.

BEST COPY AVAILABLE

households of the Head Start group had a significantly greater *number of children*. (See Table C.1 on p. 136.) At the Florida site, there were significantly more children (including siblings, cousins, and others) in the Head Start households than in the No–Head Start households (M = 4.5 [SD = 2.5] vs. M = 3.9 [SD = 2.5]). Also, at the Colorado site, there were significantly more children in the Head Start households than in the No–Head Start households (M = 3.8 [SD = 2.2] vs. M = 2.8 [SD = 1.5]).

➤ **At the Florida site, compared with the No–Head Start study participants, Head Start study participants came from original households with significantly fewer adults. At the Colorado site, there were no differences between Head Start and No–Head Start households on *total number of adults* in the original household.** (See Table C.1 on p. 136.) When the LTBHS study group participants were 5 to 6 years old, the proportions of two-parent households at the Florida site were comparable for the Head Start and No–Head Start groups: 64% of Head Start households were two-parent, compared with 67% of No–Head Start households. In Colorado, 79% of both the Head Start and the No–Head Start households were two-parent. However, in Florida, there was a significantly lower total number of adults in the Head Start households than in the No–Head Start households (M = 2.0 [SD = 0.9] vs. M = 2.2 [SD = 1.5]). In Colorado, there were no significant Head Start versus No–Head Start differences on total number of adults in household (M = 2.0 [SD = 1.3] vs. M = 2.1 [SD = 0.8]).

➤ **At the Florida site, Head Start and No–Head Start mothers did not differ significantly on *level of education*. At the Colorado site, compared with No–Head Start mothers, Head Start mothers had significantly lower *levels of education:* Significantly fewer had at least attained *high school graduation or the GED*, and significantly fewer had at least completed *some postsecondary education*.** (See Table C.2 on p. 136 for frequencies including missing data; see Marcus's Appendix F, Tables F.1 and F.2, for frequencies without the missing data and for significance test results.) At the Florida site, comparable proportions of Head Start and No–Head Start study participants—21% and 16%, respectively, did not know the highest year of education completed by their mother. From those study participants who knew their mother's highest year of education, we learned that Head Start mothers had on average completed 11.3 years (SD = 2.1); this compared with an average of 11.5 years (SD = 2.4) for No–Head Start mothers.

Thus, in the cases of Florida study participants who knew their mother's educational level, the Head Start and No–Head Start mothers had comparable educational levels: 66% of Head Start study participants reported that their mother had at least graduated from high school; this compared with 68% for No–Head Start mothers. Of Head Start mothers, 34% were reported to have less than a high school education; this compared with 32% of No–Head Start mothers. Of Head Start mothers, 16% were reported to have at least some postsecondary education; this compared with 19% of No–Head Start mothers.

At the Colorado site, similar proportions of Head Start and No–Head Start study participants (12% vs. 15%) did not know the highest year of education completed by their mother. From those who knew their mother's highest year of education, we learned that Head Start mothers averaged significantly fewer years of education than did No–Head Start mothers (M = 9.6 years [SD = 3.0] vs. M = 11.9 years [SD = 2.2]).

Thus, in the cases of Colorado study participants who knew their mother's educational level, the Head Start and No–Head Start groups differed significantly on mother's educational level: Of the Head Start group, 36% reported their mothers had at least graduated from high school, which is fewer than the 79% of No–Head Start mothers who were reported to have done so; 64% of Head Start mothers, compared with 21% of No–Head Start mothers, had less than a high school education. Of the Head Start mothers, 12% had completed at least some postsecondary education, which is significantly fewer than the 24% of No–Head Start mothers who had done so.

➤ **At both sites (Florida and Colorado), compared with No–Head Start fathers, significantly fewer Head Start fathers had achieved at least *high school graduation or GED certification.*** Similar, high proportions of Florida's Head Start and No–Head Start study participants did not know their father's educational level (38% vs. 37%). From those who knew their father's highest year of education, we learned that the Head Start fathers had completed significantly fewer years of education than the No–Head Start fathers had ($M = 10.4$ years [$SD = 2.7$] vs. $M = 11.5$ years [$SD = 2.5$]).

In the cases of Florida study participants who knew their father's educational level, significantly fewer of Head Start versus No–Head Start fathers had achieved at least high school graduation or GED certification (56% vs. 71%); 44% of Head Start fathers and 29% of No–Head Start fathers were reported to have less than a high school education. There was no significant difference in the proportion of Florida's Head Start versus No–Head Start fathers who were reported to have achieved at least some postsecondary education (12% vs. 19%).

In Colorado, more Head Start than No–Head Start study participants did not know their father's educational level (29% vs. 15%). From those who knew their father's highest year of education, we learned that the Head Start fathers had completed significantly fewer years of education than the No–Head Start fathers had ($M = 9.5$ years [$SD = 3.2$] vs. $M = 11.1$ years [$SD = 2.8$]).

In the cases of study participants who knew their father's educational level, significantly fewer of Colorado's Head Start versus No–Head Start fathers (35% vs. 60%) had at least attained high school graduation or GED certification; 65% of Head Start fathers and 40% of No–Head Start fathers had less than a high school education. Significantly fewer Head Start fathers versus No–Head Start fathers had achieved at least some postsecondary education (8% vs. 19%).

➤ **In Colorado, compared with No–Head Start mothers and fathers, Head Start mothers and fathers had *job types* that were rated lower on Duncan's Socioeconomic Index (SEI).** (See Table C.3 on p. 137.) As explained in Chapter 4, the type of employment of the mothers and fathers of the study participants (about the time of Head Start or school entry) was determined using a combination of sources—the follow-up interview, the family contact interview, or the school enrollment records. In Florida, from these sources, we were able to determine job types for 53% of Head Start mothers versus 47% of No–Head Start mothers. If we considered only those mothers who were reported to have had jobs at the time, we were able to determine job types for 63% of Head Start mothers and 65% of No–Head Start mothers. For fathers, we were able to determine job types for 62% of the Head Start group and 66% of the No–Head Start group. (Only 1 Head Start father and 2 No–Head Start fathers reportedly had no job at the time.) There were no significant Head Start versus No–Head Start differences in Duncan's SEI job scores, either for mothers or for fathers.

At the Colorado site, we were able to determine job types for 43% of Head Start mothers versus 38% of No–Head Start mothers. If we consider only those mothers who were reported to have had jobs at the time, we were able to determine job types for 57% of Head Start mothers versus 59% of No-Head Start mothers. We were able to determine job types for 80% of Head Start fathers versus 85% of No–Head Start fathers. (Only 1 of the Head Start fathers and none of the No–Head Start fathers reportedly had no job at the time.) There was a nonsignificant trend ($p < .10$) of lower SEI job scores for Head Start mothers versus No–Head Start mothers ($M = 23.4$ [$SD = 18.7$] vs. $M = 31.4$ [$SD = 22.8$]). Also, the SEI job scores of Head Start fathers were significantly lower than those of the No–Head Start fathers ($M = 15.2$ [$SD = 10.9$] vs. $M = 22.2$ [$SD = 18.9$]).

➤ **At both sites (Florida and Colorado), the Head Start group differed significantly from the No–Head Start group on the variable *family welfare assistance*.** (See Table C.4 on p. 137 and Table F.1 on p. 181 for significance test results.) In Florida, 2% of the Head Start group and 3% of the No–Head Start group (roughly comparable percentages) were unab[illegible] recall whether their families had been on we[illegible]uring their elementary school years. A significantly lower proportion of the Head Start group's families had *never* used welfare assistance during the elementary school years (46% Head Start vs. 59% No–Head Start); also a higher percentage of them reported frequent use of welfare (33% Head Start vs. 24% No–Head Start), which is a nonsignificant trend.

In Colorado, 2% of the Head Start group and 4% of the No–Head Start group were unable to recall information about family use of welfare in their elementary school years. We learned that a lower proportion of Head Start families had *never* used welfare assistance during those years (48% Head Start vs. 66% No–Head Start). However, unlike the Florida site, this site did not have a significantly greater percentage of Head Start families with frequent use of welfare.

Summing up family background—Head Start versus No–Head Start

As summarized in Table 5.3 on p. 73, in Florida the Head Start group, when compared with the No–Head Start group, showed a pattern of families having lower socioeconomic status: For Head Start families, there were fewer adults *and more* siblings *and* children per household, *fathers had less* education, *and there was a greater proportion of families who used* welfare assistance *as well as a greater* frequency of welfare use.

In Colorado, the Head Start versus No–Head Start group differences were somewhat more marked. Head Start households had lower socioeconomic status, as indicated by the following: There were more children *and more* siblings per household. *Mothers and fathers both had less* education, *and they both had lower* job status *as well. In addition, the proportion of families who had used* welfare assistance *was greater.*

Colorado's more pronounced study group differences are likely to be related to the racial/ethnic imbalance between the Head Start group and the No–Head Start group—the fact that there were significantly more Hispanic Americans in the former. Thus, we expected to encounter more difficulty at the Colorado site when adjusting for differences in estimating the long-term benefits of Head Start.

Overall, although the background comparability of the Head Start and No–Head Start groups is considerable, they appear to have been selected from somewhat different populations, with both Head Start and No–Head Start study participants having low socioeconomic status but the Head Start subjects coming from households of more limited economic means. Our analyses of core outcome variables would thus require utilizing the covariates to provide some adjustment for these family-background differences between the groups.

Comparing the Model Head Start group and the Regular Head Start group at the Florida site

➤ **There were no significant differences between Model and Regular Head Start groups on *number of children* in the original household or on *number of siblings*.** The Model Head Start group and the Regular Head Start group had families with comparable numbers of siblings ($M = 4.4$ [$SD = 2.7$] vs. $M = 4.0$ [$SD = 2.6$]) and comparable numbers of children ($M = 4.8$ [$SD = 2.6$] vs. $M = 4.3$ [$SD = 2.4$]).

➤ **There was no significant difference between the Model and Regular Head Start groups on *number of adults* in the original household.** When the study participants were 4 to 6 years old, the number of adults in the original household was the same for the Model and Regular Head Start groups ($M = 2.0$ [$SD = 0.9$]). The two study groups also had the same proportion of two-parent households (64%).

➤ **The Model and Regular Head Start groups did not differ significantly on *highest year of education completed* by mother, on proportions of mothers achieving at least *high school graduation or GED completion,* or on proportions of mothers achieving at least *some postsecondary education.*** The proportions of study participants unable to recall their mother's educational level were 30% of the Model Head Start group and 14% of the Regular Head Start group. From those who could recall this information, we learned that there was no significant difference between Model and Regular Head Start mothers in highest year of education completed (*M* = 11.0 years [*SD* = 2.0] vs. *M* = 11.5 years [*SD* = 2.2]).

If we consider only those mothers whose educational level was reported to us, a lower proportion of Model Head Start mothers than Regular Head Start mothers had completed at least high school or the GED (60% vs. 70%). However, the difference on high school/GED completion was not a significant difference. The proportion of mothers achieving at least some postsecondary education was lower for the Model Head Start group than for the Regular Head Start group (12% vs. 19%), also not a significant difference.

➤ **Model versus Regular Head Start fathers appeared to have a slightly lower level of education. There were no significant differences, however, on *highest year of education completed,* on proportions achieving at least *high school graduation/GED completion,* or on proportions achieving at least *some postsecondary education.*** Comparable percentages of Model and Regular Head Start participants were unable to recall their father's level of education (38%). In comparing highest year of education completed, we did find a nonsignificant trend ($p < .10$) of a lower level of education among Model Head Start fathers (*M* = 9.9 years [*SD* = 2.8] vs. *M* = 10.9 years [*SD* = 2.6]).

From those who knew their father's educational level, we learned that a higher percentage of Model Head Start fathers had *not* achieved high school graduation or GED completion (52% vs. 38%, not a significant difference). Comparable percentages of the Model and Regular Head Start fathers had achieved at least some postsecondary education (11% vs. 13%, respectively, not a significant difference).

➤ **The Model Head Start group and the Regular Head Start group had no significant differences on *mother's* and *father's job type* as indicated by Duncan's SEI ratings.** From the various sources available to us, we were able to determine 51% of the Model Head Start mothers' job types (if any) compared with 57% of the Regular Head Start mothers' job types. If we consider just those mothers who reportedly had jobs at the time, we were able to determine the job types for the Model and Regular Head Start mothers at comparable levels (58% vs. 66%, respectively). Comparable proportions of fathers' job types could also be determined—59% for the Model Head Start fathers, 64% for the Regular Head Start fathers. (None of the Model Head Start fathers and just 1 of the Regular Head Start fathers reportedly had no job.) In comparing the two study groups, we found no significant differences on SEI job ratings for mothers (*M* = 21.5 [*SD* = 16.9] for Model Head Start; *M* = 21.0 [*SD* = 17.8] for Regular Head Start) or for fathers (*M* = 20.0 [*SD* = 17.0] for Model Head Start; *M* = 17.9 [*SD* = 15.2] for Regular Head Start).

➤ **The Model Head Start group and Regular Head Start group had no significant differences on the variable *family welfare assistance.*** Only 3 Model Head Start study participants and none of the Regular Head Start study participants did not know whether their family had used welfare during the elementary school years. Forty-three percent of the Model Head Start group versus 48% of the Regular Head Start group reported that their families *never* used welfare assistance when the study participants were in elementary school (not a significant difference). Frequent use of welfare assistance was reported by comparable proportions of the Model and Regular Head Start participants (32% vs. 33%).

Summing up family background—Model Head Start versus Regular Head Start

For the Model versus Regular Head Start groups in Florida, there was no pattern of significant family-background differences. We found just one non-significant trend of Model Head Start fathers' having achieved a lower level of education.

The Site Tract-Level Characteristics of Study Participants' Original Households

Comparing the Head Start groups and the No–Head Start groups at the two sites

Further insights on the family backgrounds of study participants were obtained by examining potential differences in poverty levels between sites, within sites, and between the study groups within sites. As explained in Chapter 3, p. 28, the LTBHS Head Start study group in Florida included the model and regular Head Start groups from an earlier study, the Head Start Planned Variation study. To these original groups, because the HSPV study's model Head Start group had contained both rural and urban subjects but its regular Head Start group had not, the LTBHS researchers added a group of rural Regular Head Start participants and a group of rural No–Head Start participants, so the LTBHS study would have both rural and urban subjects in its Regular Head Start and No–Head Start groups as well as in its Model Head Start group. Thus we could consider the four different "areas" of the LTBHS Florida site that are listed in Table 5.4.

Data analyst Yu Xie first utilized the Geographic Information System (GIS) database and plotted on maps all the study participants' original household addresses (see Appendix D for more details). Then, U.S. Bureau of the Census tract-level statistics were derived based on the residence of each study participant. These data were based on 1970 U.S. Bureau of the Census information in all but one county of the Florida site. In this one Florida county, the 1970 population density was not sufficient to require tract-level data, so 1980 U.S. Bureau of the Census tract-level data were used. Overall, in the various areas of the Florida site, there was a substantial degree of correlation between 1970 and 1980 census data on the following socioeconomic indicators: poverty rate, $r = .87$; high school dropout rate, $r = .42$; and female-headed household rate, $r = .83$.

➤ **Compared with the No–Head Start study participants, the Head Start study participants were geographically more concentrated, and some of the areas they lived in were more impoverished.** The maps drawn up by Xie demonstrated that, compared with the No–Head Start study participants, the Head Start study participants were more concentrated geographically (see, for example, the map on p. 138 in Appendix C). The derived tract-level statistics are summarized in Table 5.4

An examination of Table 5.4 poverty rates shows that the Colorado study groups lived in an area with a poverty rate comparable to that of one of the rural Florida areas (the "rural Regular Head Start" area). Also, the Colorado site was considerably less impoverished than one of the urban Florida areas (the "urban Regular Head Start" area), which had the most impoverished Head Start and No–Head Start groups of all. Overall across all the site areas, the Head Start study groups, compared with the No–Head Start study groups, appear to be concentrated in areas with comparable or higher poverty rates.

Summing up site tract-level characteristics—Head Start versus No–Head Start

Comparison of the site tract-level characteristics *revealed some differences in original levels of poverty for study group participants who had resided in various areas of the Florida site and the Colorado site. Across the Florida and the Colorado sites, the No–Head Start groups appeared to have lived in either comparable or somewhat less impoverished*

Table 5.4: U.S. Bureau of the Census Tract-Level Data on Areas Within the Two Study Sites

Site/Study-group Area	Group	N	Site Tract-Level Characteristics					
			Poverty Rate		High School Dropout Rate		Female-Headed Household Rate	
			Mean %	SD	Mean %	SD	Mean %	SD
Florida urban Regular Head Start	Regular Head Start	71	32.1	14.1	20.2	8.5	24.6	10.1
	No-Head Start	195	34.8	12.5	23.2	7.6	26.2	9.7
Florida urban Model Head Start	Model Head Start	39	15.7	10.6	11.2	6.8	19.6	12.5
	No-Head Start	9	7.4	2.7	9.2	3.7	16.2	5.5
Florida rural Regular Head Start	Regular Head Start	17	20.1	8.6	20.6	8.1	10.4	3.4
	No-Head Start	24	19.3	8.5	17.8	8.0	9.3	4.2
Florida rural Model Head Start	Model Head Start	35	31.0	10.8	15.5	7.5	30.1	12.1
	No-Head Start	36	23.9	12.3	16.4	9.4	22.1	14.2
Colorado	Head Start	127	21.2	7.4	17.1	9.7	18.1	12.6
	No-Head Start	72	18.2	6.9	14.7	8.2	14.4	10.9

Note. These data (N = 625) include 3 study participants who were later eliminated from the final study sample (see Chapter 3). Mean % is the average percentage across U.S. Bureau of the Census tracts within each group and site; SD = standard deviation. Statistics are based on the 1970 U.S. Bureau of the Census data, except for urban and rural Model Head Start areas in Florida that lacked tract-level data in 1970; statistics for these areas were based on 1980 U.S. Bureau of the Census data.

areas than did the Head Start groups. Therefore, in analyzing the LTBHS study's outcomes data, we would need to make statistical adjustments for these initial group-differences.

Comparing the Model Head Start group and the Regular Head Start group at the Florida site

➤ **Florida's Model Head Start and Regular Head Start areas had differing economic conditions.** The data presented in Table 5.4 reveal that the urban Model Head Start area had a *lower* rate of poverty than did the urban Regular Head Start area (16% vs. 32%). However, the rural Model Head Start area had a *higher* rate of poverty than did the rural Regular Head Start area (31% vs. 20%).

Summing up site tract-level characteristics—Model Head Start versus Regular Head Start

The Model and Regular Head Start groups in Florida differed in their levels of poverty, depending on the poverty rate of the particular urban or rural area where each of these Head Start groups was concentrated. Within the Florida Head Start groups, the urban Model Head Start study participants resided in an area that was less impoverished than the area of the urban Regular Head Start study participants, while the reverse appeared to be the case for the areas of residence of the rural Model and rural Regular Head Start study participants. Along with the differences found between Model and Regular Head Start on the variable of kindergarten participation, *this mixed picture of the two groups'* site tract-level characteristics *points to the need for cautious interpretation of findings on their educational and life-situation outcomes.*

BASED ON BACKGROUND CHARACTERISTICS—THE EXPECTED PATTERN OF FINDINGS ON LONG-TERM BENEFITS

As described in Chapter 4 and earlier in this chapter, selection of the LTBHS study's core covariates and core outcome variables was based on the High/Scope Perry Preschool study's findings. However, whereas the preschool group and the no-preschool (control) group in the High/Scope Perry Preschool study had only one significant difference on family-background variables, the foregoing summary has revealed that the LTBHS study's Head Start and No–Head Start groups had many significant family-background differences. Furthermore, although comparison of most of the Florida Model versus Regular Head Start groups' family-background characteristics revealed no significant differences, comparison of the groups' site tract-level characteristics *did* reveal some urban/rural differences. As already noted, all the various differences that were found pointed to the need for caution in interpreting any findings on the core outcome variables of the LTBHS study.

In addition to the High/Scope Perry Preschool study, two other related studies helped to frame our expectations of LTBHS study findings—the Head Start Planned Variation (HSPV) study and the High/Scope Preschool Curriculum Comparison study, which we also examine here.

Comparison of the LTBHS Study and Related Studies: Participants' Backgrounds

The following considerations make up the rationale for comparing the LTBHS study with the three related studies—the HSPV study, the High/Scope Perry Preschool study, and the High/Scope Preschool Curriculum Comparison study:

- The studies all employed the High/Scope preschool curriculum.
- The HSPV study compared model and regular Head Start study participants from 1970–71, who are included in the LTBHS study's Florida site.
- The HSPV study looked at short-term Head Start effects when comparing its model (High/Scope) Head Start group with its regular Head Start group.

- The High/Scope Perry Preschool study compared a High/Scope preschool group to a no-preschool (control) group and collected long-term effects data on young-adult subjects.
- The High/Scope Preschool Curriculum Comparison study compared a High/Scope preschool group with two other types of preschool groups and collected data on long-term effects of the three types of preschool through young adulthood.

To the extent that these three studies are comparable to the LTBHS study in key available background factors, such as raçial/ethnic membership and socioeconomic indicators, they give further basis for expecting similar outcomes in the LTBHS study. To assess the comparability of the LTBHS study sample to those of other studies, the following discussion examines the key family- background variables that the four studies had in common.

The Head Start Planned Variation study

As discussed previously, the LTBHS study was rooted in the HSPV study in Florida and Colorado, where the HSPV study had found significant Stanford Binet IQ gains for model (High/Scope) Head Start participants compared with regular Head Start participants (Smith, 1973).

In sites across the nation, the HSPV study investigators found differences in sites and family-background characteristics for the study's various model and regular Head Start groups (as reflected in the U.S. Bureau of the Census site characteristics data in the Florida site that was carried over into the LTBHS study). As in the LTBHS study, the HSPV data analysts employed many analytic strategies to adjust for such differences.

In the HSPV study (Smith, 1973, pp. 45, 50), in the total Florida model Head Start group (*n* = 90), 69% were African American and 31% were White; in that site's total regular Head Start group (*n* = 66), 85% were African American and 15% were White. In the study's total Colorado Head Start group (*n* = 136; this is model and regular combined) 64% were Hispanic American and 36% were White/other.

For the HSPV data analysis, the socioeconomic status data from the HSPV study analyses and all subsequent analyses were based on a subsample of this total sample. (Study participants were eliminated from the original HSPV study sample for a variety of reasons: Some had not completed the total study assessment battery in the fall of 1970; some had not been assessed in both fall 1970 and spring 1971; or some did not have verified background variable codes for gender, age, other child care experience, or racial/ethnic membership.) Table 5.5 on p. 82 presents family background data for the HSPV study's final data analysis sample (Smith, 1973, pp. 59–60).

The final HSPV data analysis sample for the Florida site contained nearly 70% of the original HSPV sample. By contrast, the LTBHS study sample at the Florida site, in its Head Start group, included nearly 92% of the original total HSPV study sample as well 17 additional rural area Regular Head Start participants—from the rural Florida area adjacent to and served by the urban Head Start agency.

As shown in Table 5.5, in Florida the HSPV data analysis sample included 58 Model Head Start students (76% African American and 24% White) and 51 Regular Head Start students (90% African American and 10% White). Thus, the racial/ethnic representation was highly similar for the HSPV data analysis sample and the LTBHS Head Start groups—both studies had predominantly African Americans in the Florida Head Start groups and predominantly Hispanic Americans in the Colorado Head Start group. The HSPV Florida Regular Head Start sample did have a higher proportion of African Americans, 90%, compared with 83% in the LTBHS study sample; this is likely a result of the LTBHS study's addition of the rural-area Regular Head Start group. This addition, however, succeeded in making the LTBHS study's Model and Regular Head Start groups more comparable to

Table 5.5: Comparison of Participants' Family Backgrounds in the LTBHS Study and Related Studies

Study/Group	*N*	Race/Ethnicity	Household Density *M* persons	Mother's Education *M* years	Other Socioeconomic Indicators[a]
HSPV Study					
FL model Head Start	58	76% African Amer. 24% White	6.4	10.1	1970 Income: $3,310
FL regular Head Start	51	90% African Amer. 10% White	5.7	9.8	1970 Income: $2,890
CO Head Start (all)	76	68% Hispanic Amer. 32% White	5.4	9.6	1970 Income: $4,545
High/Scope Preschool Curriculum Comparison Study[b]					
3 preschool groups combined	68	65% African Amer. 35% White	6.7	10.0	67% not on welfare at study entry
High/Scope Perry Preschool Study					
Preschool/no-preschool groups combined	123	100% African Amer.	6.7	9.5	49% not on welfare at study entry
LTBHS Study					
FL Head Start (all)	160	78% African Amer. 22% White/other	6.5	11.3	46% never on welfare (elem. school years)
Model	74	73% African Amer. 27% White/other	6.8	11.0	43% never on welfare (elem. school years)
Regular	86	83% African Amer. 17% White/other	6.3	11.5	48% never on welfare (elem. school years)
FL No-Head Start	264	80% African Amer. 20% White/other	6.1	11.5	59% never on welfare (elem. school years)
CO Head Start	130	73% Hispanic Amer. 27% White/other	5.8	9.6	48% never on welfare (elem. school years)
CO No-Head Start	68	32% Hispanic Amer. 68% White/other	4.8	11.9	66% never on welfare (elem. school years)

Note. FL = Florida; CO = Colorado. The High/Scope Perry Preschool (Schweinhart et al., 1993) and Preschool Curriculum Comparison (Schweinhart & Weikart, 1997) studies were conducted in Ypsilanti, Michigan. HSPV = Head Start Planned Variation study; LTBHS = Long-Term Benefits of Head Start study. The HSPV study sample shown here is based on the final sample employed for the data analysis (Smith, 1973, pp. 59–60).

[a] The Federal poverty threshold in 1970 for a family of four persons was $3,800 (nonfarm) and $3,200 (farm).

[b] Although the majority of the study participants' families in the High/Scope Preschool Curriculum Comparison study did not utilize welfare, all the families were living in poverty-level households.

each other on ethnicity as well as on types of areas of residence. Still, the groups in the HSPV study's data analysis sample were comparable to the groups in the LTBHS study's sample regarding household density. Although the LTBHS mothers were roughly comparable in educational level to mothers in the HSPV data analysis sample, the LTBHS mothers appear to have more education.

The HSPV study variables regarding socioeconomic status are also included in Table 5.5. According to their 1970 income figures, the HSPV model and regular Head Start groups at the Florida site were essentially at the same economic level, and the combined (model and regular) Colorado Head Start group was at a higher level than the Florida Head Start groups. All HSPV group incomes shown in Table 5.5 were considered to be at or below the Federal poverty threshold for a family of four in 1970 (indicating eligibility for welfare assistance and Head Start)—$3,800 (nonfarm) and $3,200 (farm).

The High/Scope Preschool Curriculum Comparison study

The High/Scope Preschool Curriculum Comparison study (Schweinhart & Weikart, 1997) included a High/Scope curriculum group that was found to have gained several long-term benefits when it was compared against two groups using other preschool curriculums. Comparing the three study groups of the High/Scope Preschool Curriculum Comparison study may be analogous to comparing the Model (High/Scope) Head Start group and the Regular Head Start group at the Florida site of the LTBHS study. The High/Scope Preschool Curriculum Comparison study employed random assignment to its three curriculum groups, and there were no significant family-background differences among the three groups that were compared (see Table 5.5). The study's sample household density averaged 6.7 persons, compared with averages of 6.8 persons for the LTBHS Model Head Start group and 6.3 persons for the LTBHS Regular Head Start group. The average level of mother's education for the Curriculum Comparison study (10.0 years) was less than those for the LTBHS study's Model and Regular Head Start mothers (11.0 years and 11.5 years, respectively). The Curriculum Comparison study group's use of welfare assistance was the lowest—67% of the families were not on welfare at the time of the subject's entry into preschool (Schweinhart, Weikart, & Larner, 1986, p. 23). This percentage is most comparable to the LTBHS study's Florida No–Head Start group's 59% and Colorado No–Head Start group's 66% (both LTBHS percentages referring to lack of welfare use during the study participant's elementary school years).

The High/Scope Perry Preschool study

As indicated throughout this book, the most relevant study to compare with the LTBHS study is the High/Scope Perry Preschool study (Schweinhart et al., 1993), which also employed the High/Scope preschool curriculum in one study group and then compared that preschool group to a no-preschool group. Because study participants were randomly assigned to the preschool and no-preschool groups, no significant family-background differences were found. As shown in Table 5.5, there were 6.7 persons per original family household in the High/Scope Perry Preschool study's combined group, which was most comparable to the LTBHS study's highest household density (6.8 persons for its Florida Model Head Start group) and somewhat less comparable to the LTBHS study's lowest household density (4.8 persons for its Colorado No–Head Start group). In the High/Scope Perry Preschool study (p. 49), about 53% of the study participants originally came from two-parent households, which is less than the proportions of two-parent households in the LTBHS study's Florida Model and Regular Head Start groups (64% each), Florida No–Head Start group (67%), and Colorado Head Start and No–Head Start groups (79% each).

In the High/Scope Perry Preschool study (Schweinhart et al., 1993, p. 48), the initial parent interview at the time of study entry revealed that the preschool group's mothers completed an average of 9.5 years of schooling (*SD* = 2.4), compared with 9.4 years of schooling (*SD* = 2.0) for mothers of the no-preschool group; thus the High/Scope Perry study had no between-group difference, and it had a lower level of education for mothers than did the LTBHS study groups. In Florida, the LTBHS Model Head Start mothers averaged 11.3 years of education (*SD* = 2.1), and the No–Head Start mothers averaged 11.5 years of education (*SD* = 2.4). More comparable with the High/Scope Perry Preschool study sample, the LTBHS study's Colorado Head Start mothers averaged 9.6 years of education (*SD* = 3.0), whereas Colorado No–Head Start mothers averaged 11.9 years (*SD* = 2.2).

Looking at the "other socioeconomic indicators" in Table 5.5 reveals that at the LTBHS study's Florida site, approximately half (54%) of the Head Start participants' families *had* been on welfare at some point during the elementary school years. The percentage of families on welfare in the High/Scope Perry Preschool study (Schweinhart et al., 1993, p. 49) suggests a similar level of poverty at subjects' preschool entry. Whereas in the High/Scope Perry Preschool study there were no significant differences between the preschool and the no-preschool groups on the family-background variables shown in Table 5.5, in the LTBHS study the No–Head Start groups at both the Florida site and the Colorado site were more advantaged economically than the Head Start groups. According to the various socioeconomic indicators of Table 5.5, the LTBHS study's No–Head Start groups (especially the Colorado No–Head Start group) had a higher socioeconomic status than the comparison study groups had in these related studies.

Overall, as shown in Table 5.5 and the corresponding discussion, the study groups in these related studies were fairly comparable in family socioeconomic status in their early childhood years. The findings of these studies are thereby relevant for further consideration in formulating a set of expected LTBHS study findings, as presented next.

Expected Findings for the LTBHS Study

Given the Table 5.5 comparisons, along with earlier reported background differences between the LTBHS study's Head Start and No–Head Start groups, the following discussion presents the expected pattern for the long-term outcomes in the LTBHS study.

Predictions for the Head Start versus No–Head Start groups

As discussed in Chapter 4, the core outcome variables for the LTBHS data analyses were selected because they were highly similar to outcome variables on which researchers found long-term benefits for the preschool group in the High/Scope Perry Preschool study. However, as noted before, since the High/Scope Perry Preschool study employed randomly assigned study groups, its estimate of effects should be superior to that of the LTBHS study, which employed nonrandom study groups. Furthermore, because the No–Head Start group in the LTBHS study was originally less impoverished economically than the Head Start group, we could reasonably expect to find better life outcomes for the No–Head Start group than for the Head Start group. However, to the extent that Xie's and Marcus's analytic models effectively adjust the LTBHS study outcomes data to compensate for the original socioeconomic differences and other types of background differences between the Head Start and No–Head Start groups, we could expect to find Head Start effects that are comparable at least in direction if not also in magnitude to the preschool effects found in the High/Scope Perry Preschool study. To the extent that their analyses do not fully adjust for the background differences of the study groups, we could expect the LTBHS study to arrive at an underestimation of Head Start benefits.

If the LTBHS study finds any indications of Head Start long-term benefits, future studies of Head Start effectiveness (concentrating on Head Start programs of sufficient program quality) should try to either employ a random-assignment design or conduct a more complete assessment of their study groups' background differences. In such studies, we should find more evidence of long-term benefits for Head Start than we could expect to find in the LTBHS study.

Because, on educational outcomes, the High/Scope Perry Preschool study's long-term benefit findings were stronger for females than for males, we also could expect the LTBHS study to find that Head Start and/or Model Head Start has greater educational benefits for females than for males.

Predictions for the Model versus Regular Head Start groups

Regarding expectations for the LTBHS study's Model versus Regular Head Start comparison, the urban Model Head Start area at the Florida site (see Table 5.4 on p. 79) could be expected to provide a less detrimental environment than that provided by the urban Regular Head Start area. Because of a possible boost from having participated in a Head Start program using High/Scope's child development, theory-based curriculum (as suggested by findings from the High/Scope Preschool Curriculum Comparison study and the HSPV findings) as well as the possible boost from living in a somewhat better-off community, we could expect to find some benefits for the Model Head Start group over and above those for the Regular Head Start group.

Since Florida's rural Model Head Start area appeared to be more disadvantaged economically than its rural Regular Head Start area, we could expect the rural Regular Head Start group to do better as young adults than the rural Model Head Start group, unless experiencing the Model Head Start program were to have a sufficiently compensatory effect on children. Unfortunately, the statistical analytical models necessitated combining the urban and rural areas within both the Model and the Regular Head Start study group, resulting in mixed levels of economic impoverishment within each of those two groups at the Florida site. This precluded our having clear expectations and limited our later interpretations of findings. We do know that the Model and Regular Head Start groups had comparable family background characteristics. Yet, of the two groups, the Model Head Start group had the greater kindergarten participation.

If Model Head Start study participants, when compared with Regular Head Start participants, are found to have better educational or life-situation outcomes, it could suggest that (1) the High/Scope curriculum (or a similar child development, theory-based curriculum) being used in a Head Start program can make a lasting contribution or (2) other factors, possibly related to site characteristics (including greater opportunities for kindergarten), contribute to better outcomes. This latter interpretation would suggest that Head Start may be more likely to provide long-term benefits when there are supportive educational or employment opportunities in the community. If no advantage is found for the Model Head Start group, this would suggest that using a model curriculum or living in a community with greater opportunities does not provide any opportunity for Head Start participants, over the long run, to improve their economic status.

A SUMMARY OF EXPECTATIONS BASED ON BACKGROUND COMPARISONS

First, given the findings of long-term benefits in the High/Scope Perry Preschool study, we expected that to the extent that the analytic models employed by Xie and Marcus adjusted for the obtained background differences in the LTBHS study, we would find some indications of long-term benefits for Head Start participation. Considering the study biases inherent in the nonrandom design, we expected that to the extent that the analytic models did not make sufficient adjustments with the available covariates, we would find the No–Head Start group to be in the more advantaged situations as young adults. Furthermore, in Colorado, given the markedly lower socioeconomic status of the Head Start group versus the No–Head Start group, we expected that statistical adjustment for these factors might not be sufficient to reveal long-term benefits of Head Start. Also in Colorado, the greater proportion of Hispanic Americans versus Whites/others in the Head start group, when compared with the No–Head Start group, might constitute an advantage for the No–Head Start group in that its mem-

BEST COPY AVAILABLE

bers might experience less discrimination in seeking school, employment, and various other kinds of opportunities.

Second, we expected that better long-term outcomes for the Model (High/Scope) Head Start group would indicate some plausible advantage for using a child development, theory-based curriculum in Head Start programs. Such a finding would parallel the promising findings from the HSPV study, the High/Scope Perry Preschool study, and the High/Scope Preschool Curriculum Comparison study. However, if the analytic models cannot control for the Model versus Regular Head Start groups' apparent variations in site tract-level characteristics, any indication of Model Head Start effects would be only suggestive, rather than conclusive, regarding the advantage of using a model curriculum.

Third, because the High/Scope Perry Preschool study found some advantage for females in its preschool group, we expected that the LTBHS study analyses by gender would yield stronger effects for females who participated in Head Start or in the Model (High/Scope) Head Start. The three studies related to the LTBHS study provide no empirical basis for predictions regarding any differential findings on racial/ethnic membership or socioeconomic status.

6

MAJOR FINDINGS ON HEAD START'S LONG-TERM BENEFITS

The Long-Term Benefits of Head Start (LTBHS) study found that some groups of study participants were better-off as a result of their Head Start participation. One finding was that the Florida site females who had attended Head Start reached higher educational levels than they otherwise would have. We also found a pattern indicating potential benefits for the Model (High/Scope) Head Start group versus the Regular Head Start group (the Model Head Start group was carried over from the 1970 Head Start Planned Variation [HSVP] study). However, the site-characteristic differences made it difficult to fully interpret this latter finding. Statistical analyses of study-group comparabilitiy indicated initial family-background and study-site differences between the LTBHS study's Head Start and No–Head Start study groups. When a degree of statistical adjustment for these initial background differences was made, the likelihood of Head Start benefits increased. Before these differences were taken into account, the result was an underestimation of the long-term effects of Head Start.

The findings of the LTBHS study provide evidence, but not confirmation, that Head Start participation can result in long-term benefits extending into adulthood. The evidence suggests that this may particularly be the case when the Head Start program employs a child development, theory-based curriculum. It cannot be concluded, from the data of this study, that Head Start is ineffective in providing long-term benefits for its participants. Rather, the LTBHS study findings suggest that a long-term study of Head Start effects that employs random assignment to study groups or conducts a more complete assessment of family background may find even stronger evidence of long-term benefits than this study has found. Furthermore, as a study that did not employ random assignment but that was instead quasi-experimental, the LTBHS study provides examples of several effective data analysis approaches that further quasi-experimental studies might well use to assess effects of Head Start or other educational or social service programs.

The core outcome variables employed by the Long-Term Benefits of Head Start (LTBHS) study were selected to provide a view of the study participants' pathways into young adulthood. (The covariates are listed in Table 4.1 on p. 53.)

Based on these outcome variables, this chapter begins with a description of the LTBHS study participants as young adults, at the time of the study interview (at age-22, on average). This is followed by a discussion of the effectiveness of the strategies employed by the LTBHS study's analysts to adjust for study-group background differences and thus estimate the benefits of Head Start in terms of these outcome variables. The chapter then presents the analysts' findings.

THE STUDY PARTICIPANTS AS YOUNG ADULTS

The young-adult outcomes of the 622 study participants in the LTBHS study (Head Start and No–Head Start groups combined) were these: By the time of the study interview and subsequent records-data collection, 82% of them had graduated from high school or completed the GED (the study participants were, on average, approximately 22 years old at the time of the interview, and 1990 was the year of the records-data collection). This percentage matches the 1990 national rate of high school/GED completion for 18- to 24-year-olds, which is 82% (U.S. Bureau of the Census, 1994). By the time they were young adults, 43% of the study participants had obtained some postsecondary education; the 1990 national rate for 14- to 24-year-olds being enrolled in or completing a year or more of postsecondary education is 59% (American Council on Education, 1995).

At the time of their follow-up interview, 70% of the LTBHS study's participants were either employed, enrolled in school, or both. According to interview and records data, 62% of all the study participants had *never* used welfare; 78% had *not* been a teen parent; 59% had *never* been arrested; and 70% had *never* been convicted of a crime. These statistics, which favorably describe the life situations of the majority of study participants, also indicate that for another fairly sizable proportion, the transition to adult economic independence was problematic. This is especially reflected in the proportions who *had* utilized public assistance or *had* been involved in criminal activity.

Looking beyond this overall picture of this sample of 622 young people, all of whom came from the same low-income neighborhoods and went to the same elementary schools in the early 1970s, the data analysts addressed these central questions in their analysis of the data on the Head Start and No–Head Start groups within this sample:

- *Does Head Start participation increase the likelihood that a young person will attain higher levels of education and economic viability and other positive life outcomes? Specifically, what are the long-term benefits of having participated in Head Start?*
- *Are there benefits from having participated in a Model versus a Regular Head Start program?*
- *Are there differential effects of Head Start participation according to gender, to racial/ethnic membership, or to socioeconomic status?*

ESTIMATING HEAD START BENEFITS: THE EFFECTIVENESS OF ADJUSTING FOR BACKGROUND DIFFERENCES

Working with different statistical models, data analysts Sue Marcus and Yu Xie took steps to adjust for the initial background differences of the Head Start and No–Head Start study groups to, in effect, level the playing field for the Head Start versus No–Head Start group comparisons on the core outcome variables. (Table 5.2 on p. 71 and Table 5.3 on p. 73 provide a summary of the LTBHS study groups' background differences.)

Across the various analyses, each analyst omitted one or two covariates on which no differences were found or which lacked appropriate measurement dimensions for a particular analysis approach. To lessen the variance due to site-characteristic differences, both Xie and Marcus analyzed the Florida site's data separately from the Colorado site's data. Using the data from the Florida site, the analysts also compared the two Head Start subgroups that had originated with the Head Start Planned Variation study—the Model (High/Scope) Head Start group and the Regular Head Start group. The following discussion examines the pattern of obtained results and assesses the effectiveness of the analytic models in adjusting for the Head Start and No–Head Start groups' initial differences.

As demonstrated in the following examination, the first-stage data analyses, which present the unadjusted long-term outcomes data, were found to be inappropriate for assessing Head Start's long-term benefits. These data analyses were conducted and presented by Xie and Marcus, however, to provide a base line for the later analytic models, which were to include the covariates to adjust for socioeconomic and other background differences between the Head Start and No–Head Start groups. As we will go on to describe, when their analysis included the covariates, the change in the pattern of findings on outcome variables was in a direction favorable to the Head Start group.

Adjusting for Measured Background Differences

As just noted, the analysts' first-stage analyses set forth what the outcomes data would look like without adjusting for socioeconomic background differences (see Table D.1, p. 140; Table D.14, p. 166; Table E.1, p. 181; and Table E.2, p. 182). As discussed at the end of Chapter 5 regarding the expected findings, given that the No–Head Start group began their lives in circumstances less impoverished than those of the Head Start group, we expected the No–Head Start group to be in less disadvantaged circumstances as young adults as well. Indeed, the analysts' first-stage analyses did reveal that, when compared with No–Head Start study participants, Head Start study participants as young adults had lower status educationally and in other respects.

However, each analyst cautions against making any interpretations about Head Start effects based on these unadjusted data. Rather, their data analysis challenge was this: Is it possible to estimate the extent of the No–Head Start study participants' initial advantage over the Head Start participants and effectively adjust the data to thereby lessen this advantage? If so, this would provide a more valid point of comparison for the Head Start group—an estimate of where they would be in their early adult years had they not participated in Head Start.

Because socioeconomic status is usually found to correlate positively with achievement and negatively with problem behavior or welfare dependency, failing to control for the socioeconomic background differences found between the study groups would likely result in downward (or negative) bias in simple estimates of Head Start effects (Campbell & Boruch, 1975). This supports the fact that in their first-stage, unadjusted-data analyses, both Xie and Marcus found outcome-variable differences between the Head Start and No–Head Start groups that favored the No–Head Start subjects. However, each analyst demonstrated that data analyses adjusting for the observed covariates (which had shown the significant initial socioeconomic differences between groups) subsequently narrowed, eliminated, or sometimes even reversed the Head Start versus No–Head Start differences that were originally found on educational or other life-situation outcomes. Thus, making such adjustments provided a better estimation, although still a possible underestimation, of Head Start effects. To expand on this point, we next examine the analytic approaches employed and the effects these approaches had on the outcome variables.

As described earlier, Xie generated U.S. Bureau of the Census tract-level characteristic data for each study site, based on study participants' original-household addresses. (See again Table 5.4 on p. 79, which shows that, across the two study sites, the residential areas of the Head Start groups were more impoverished than were those of the No–Head Start groups.) In his statistical analysis, Xie included the *site tract-level characteristic* covariate along with the following core covariates that showed significant Head Start versus No–Head Start differences and that could be employed effectively in his analytic models: *gender, race/ethnicity, mother's/father's educational level, family's welfare use.* Xie's Model 1 was the first-stage analysis of the unadjusted data. His Model 2 employed the battery of covariates to adjust for the measured background differences between groups, using logit regression modeling and linear regression modeling via ordinary least squares. His Model 3 also looked at the interactions of the Head Start and No–Head Start study groups with *gender* and other subgroups, to be discussed later in this chapter.

Xie's Model 2 demonstrated that whereas in the Model 1 analysis (the first-stage data analysis), most outcome variables had shown Head Start participants to be of lower status than No–Head Start participants, in the Model 2 analysis the Head Start groups shifted substantially upward on most outcome variables. For example, on *high school/GED completion,* the Model 2 data adjustment for socioeconomic status at the Florida site resulted in shifting the statistical regression coefficient in a positive direction. Although not large enough to be statistically significant, this shift gives support to the hypothesis concerning underestimation of Head Start benefits in this study.

Xie's Model 2 findings for the Colorado site data on *high school/GED completion* provide an illustration of the limitations of the statistical models. After Xie adjusted for socioeconomic differences in the study groups, the statistical regression coefficient shifted from being negative to being close to zero, which would suggest that Head Start had no measured benefit for the study participants' *high school/GED completion.* However, as noted previously, in Colorado significant racial/ethnic differences were found between the study groups (with a greater percentage of Hispanic Americans in the Head Start group). The LTBHS study data analyses found that the odds of completing high school were significantly lower for Hispanic Americans than for Whites/others. Thus, it may be that data analysis models could not fully adjust for possible study-group imbalance on some of the unmeasured factors in this study—hidden factors that influence high school completion and are also correlated with minority-group status.

Marcus's first-stage analysis also found the Head Start group versus the No–Head Start group to have lower status on educational outcome variables and other life-situation outcome variables (reported in Tables F.3, F.4, F.6, and F.7 in Appendix F). Her subsequent analysis calculated propensity scores—the probability of being in Head Start, given a battery of measured covariates—to use in matching, subclassification, and covariance-adjusted analysis, thereby adjusting for the initial background differences between Head Start and No–Head Start study participants. Marcus's battery of socioeconomic factors did not include the U.S. Bureau of the Census *site tract-level characteristics.* Nevertheless, her subsequent, adjusted-data analysis found effects similar to those found in Xie's Model 2 analysis—for most variables, the likelihood of a Head Start benefit moved from a negative to a zero effect, including a lessening of a No–Head Start advantage on *Adult APL Survey score.* After the adjustment, she also found one positive effect of Head Start participation on *elementary school GPA.*

Estimating Unmeasured Bias

In addition to adjusting for the LTBHS study's measured socioeconomic background factors, Xie and Marcus conducted analyses to estimate the probability of influential unmeasured factors in the study data. Though we have evidence that the Head Start group initially was more impoverished than the No–Head Start group, we do not know the degree of this difference and thus have limited or no ability to adjust for it. Having some estimate of how much variance is unaccounted for would improve the validity of the study's estimates of Head Start benefits.

Looking at the potential of unmeasured bias includes considering other variables in addition to those assessed in the LTBHS study. This topic of unmeasured bias was discussed in some detail in Chapter 3 on p. 31. For example, family income was not measured in the LTBHS study and thus could not be adjusted for in the data analysis. Further, parents who enrolled their children in Head Start, versus those who did not, may have held more positive, encouraging attitudes about education, and this kind of potentially influencing factor also was not measured in the LTBHS study. Similarly, the Head Start staff, in recruiting children, may have employed some selection criteria that remain unknown or unexpressed.

Based on a switching regression analysis, Xie concluded that with respect to the study participants' *highest year of education* achieved by young adulthood, the analysis did not detect any sensitivity or potential hidden selection-bias involving who had or had not enrolled in Head Start (see Table E.1 on p. 173). This may be because there *was* very little selection sensitivity or because the hidden positive sources of bias (e.g., parental encouragement of education) and hidden negative sources of bias (e.g., low economic status) offset each other. For *grade point average* (GPA) at each of three levels of schooling, Xie's analysis suggests that there was some indication of negative selection into Head Start or positive selection out of Head Start—that children who had participated in Head Start, versus those who had not, appeared to be initially less successful academically. This may, in part, suggest Head Start children's initial lower skills or capability. Regarding the variable *Adult APL Survey score*, in the Florida data there appeared to be some evidence of greater benefits from Head Start participation for African American study participants than for Whites/others, although this was not statistically significant. In total, Xie concluded that there appears to be some evidence for the hypothesis that the Head Start group, compared with the No–Head Start group, may have had additional, unmeasured disadvantages related to the assessed long-term outcomes.

To look at potential hidden bias, Marcus employed propensity scores analyses using matching and odds ratios (see Table F.8 on p. 193 and Table F.9 on p. 194). As does Xie's analysis, Marcus's analysis and interpretations give further support to the hypothesis that a more complete socioeconomic status assessment would result in better estimates of Head Start's influence. Her analysis shows that hidden bias, such as parent involvement in children's education and parental expectations, in this nonrandom design may have had some influence on several outcome variables.

The results of Marcus's analysis provide some insights on how hidden bias may affect results of potential Head Start benefits: For the Head Start versus No–Head Start comparison at the Florida site, the outcomes data were found to be sensitive to small biases, which means that small unmeasured biases could be affecting the observed outcomes. A relatively small imbalance in the study groups—due to some unmeasured covariate—could have contributed to the observed young-adult status of the Head Start and No–Head Start groups. For example, if an extra year of education (a moderate effect) could be attributed to unmeasured variables, then it is plausible that Head Start could increase the average for study participant's *highest year of education* by about ½ year. Thus, a large hidden bias could be consistent with a significant increase in *highest year of education* for the Head Start group. The existence of a moderately small hidden bias would be consistent with a beneficial effect for the Florida Head Start study group on *GPA* at various levels of schooling.

Although no significant difference in favor of Head Start was found for *GPA* in Colorado, Marcus estimated that if the impact of an unmeasured source of bias were moderately large, the effect of Head Start participation could be to raise *GPA* by about ¼ point. However, she found Colorado's Head Start versus No–Head Start comparisons on other outcomes to be relatively insensitive to moderately large hidden bias, perhaps because the two study groups appear to be very different initially. This suggests that while it is possible that Head Start is beneficial at the Colorado site, any imbalance between the groups that is not accounted for by the observed covariates would have to be fairly large to influence findings on these outcome variables. All of this contrasts with her conclusions about the Florida Head Start effects, where even a small hidden-bias factor, if determined and adjusted for, could shift a minimal or zero effect of Head Start to a benefit.

Finally, both analysts examined the potential role of missing data ("unknown" responses), which could be another source of bias influencing the estimation of Head Start effects. Using propensity scores, Marcus compared study participants who had missing data for a given outcome variable with other study participants who did not. At the Colorado site, on *GPA* in the elementary and middle school years, this analysis indicated a possible bias due to missing data. Those study participants who had missing data on *GPA* were also found to be significantly higher on socioeconomic status initially. Because a somewhat smaller proportion of the Head Start group than of the No–Head Start group had missing data, the result could be a bias in favor of the No–Head Start group, which appeared higher on *GPA*. Therefore, Head Start participation might actually be slightly more beneficial on *GPA* than is indicated by the outcome data.

To figure in the potential bias from missing data, Xie entered missing data as a category for *mother's* and *father's education* and *family's welfare use*. He also elected not to include *mother's* and *father's employment* (SEI job scores) as covariates because of the high percentages of missing data. Xie also elected not to conduct analyses on two core outcome variables—*teen-parent status* and *employment/school enrollment* at the time of the follow-up interview—because becoming a parent, being employed, or being enrolled in school can change very quickly in the transition from the teenage years to the young-adult years.

Summary of the Adjustment/Estimation Process

Employing several statistical approaches, the analysts examined the extent to which the LTBHS study's outcome variables for Head Start versus No–Head Start were affected by the groups' initial differences on the available covariates. Their analyses indicated that the long-term benefits of Head Start may be underestimated in the LTBHS study. The selection modeling analysis conducted by Xie and the hidden-bias analysis conducted by Marcus point to the possibility of additional potential sources of bias inherent in Head Start enrollment or lack of enrollment that were not measured and thus could not be employed in the adjusted data analyses. In sum, adjustment for the No–Head Start group's initial higher socioeconomic status was not sufficient to provide the best estimate of what the young-adult status of the Head Start subjects would have been had they not participated in Head Start.

The LTBHS study's limitations are due largely to (1) the limited number of valid socioeconomic status measures that were employed; (2) the initial site tract-level characteristic *differences between the Head Start and No–Head Start groups; (3) the initial socioeconomic differences between Head Start and No–Head Start groups; (4) the dependence on young-adult study participants' recall about family status in early childhood, which resulted in measurement error; (5) the Colorado groups' significant differences in* race/ethnicity *and (6) the additional unmeasured sources of bias in the selection into Head Start or No–Head Start groups. Despite the expected and demonstrated limitations of this study, the statistical models employed by the analysts provide some insights about plausible long-term benefits of Head Start, as described in the remainder of this chapter.*

THE FINDINGS ON THE CORE OUTCOME VARIABLES

The pages that follow summarize the major findings of Xie and Marcus on the outcome variables of this study. In Appendixes D, E, and F the analysts themselves fully describe all their statistical models and analyses, obtained findings, and interpretations of the findings. As in the previous chapter, throughout this chapter's description of findings, percentages have been rounded to whole numbers when appropriate. Statements of significant differences between groups indicate statistical significance with $p < .05$, although both Xie and Marcus, in their reports, also note significant findings with $p < .10$. Specific levels of significance are reported in the table notes and in the analysts' own reports.

The findings outlined here address the central research questions listed on p. 88 and the expected findings, described and discussed earlier: Head Start's plausible benefits as reflected in the Head Start versus No–Head Start group analyses at both sites; the Model Head Start versus Regular Head Start group analysis in Florida; and the major subgroup analyses—for *gender, race/ethnicity,* and *level of poverty.*

Analysis of Head Start Versus No–Head Start

As explained earlier, both Xie and Marcus conducted first-stage analyses that used unadjusted data (see Table D.1, p. 140; Table D.14, p. 166; Table F.3, p. 184; Table F.4, p. 185; Table F.6, p. 189; and Table F.7, p. 190). However, their findings from analyses that adjusted for the initial socioeconomic differences provide a better estimate of potential Head Start benefits. Throughout Xie's and Marcus's various analyses, obtaining a lower or negative Head Start coefficient is generally taken to indicate a lack of benefit for the Head Start group, which in turn indicates that the particular statistical model being employed was not able to overcome one or another of the background advantages of the No–Head Start group over the Head Start group (Campbell & Erlebacher, 1970; Manski, 1995). A positive Head Start coefficient is generally taken to indicate a benefit for the Head Start group (e.g., a benefit regarding the outcome variable *high school/GED completion*). Outcome variables such as *use of public assistance, arrest incidence,* or *conviction incidence* are exceptions, where a lower or negative Head Start coefficient indicates a benefit of Head Start participation (i.e., that Head Start reduces use or incidence).

Educational outcomes

➤ On *elementary school GPA,* some analyses indicated a benefit for Head Start at the Florida site. The first-stage, unadjusted-data analyses by both analysts found no significant benefit for the Florida Head Start group on average GPA in the elementary, middle, and high school years. The first-stage analyses of both analysts did find significantly higher GPA for the Colorado No–Head Start study participants in elementary, middle, and high school. Somewhat different findings emerged from the two analysts' subsequent adjusted-data analyses on the GPA data.

In Xie's subsequent analysis, linear regressions were estimated via ordinary least squares (OLS), and this approach found the Head Start coefficient for GPA to indicate no significant Head Start versus No–Head Start difference at the Florida site. However, his linear regression analyses did find a significant negative coefficient (indicating a lower GPA) for the Colorado Head Start group, which Xie attributed to the Colorado study groups' initial ethnic/racial imbalance. Also, Xie did identify in his adjusted analysis an effect for Head Start females on high school GPA, which will be discussed later in this chapter.

In one of Marcus's analyses that employed the core covariates to adjust for the Florida study groups' initial socioeconomic differences, she found that the benefit on elementary school GPA became significant in favor of Head Start participation, and she found a trend for the same finding in another analysis. As in Xie's analysis, no differences were found indicating a Head Start benefit on middle or high school GPA. Marcus's first-stage, unadjusted-data analyses showed a significant effect for the Colorado No–Head Start group on GPA across the grade levels (as Xie's analysis did). However, in Marcus's adjusted-data analysis, this GPA difference shifted to nonsignificance; an exception was her subclassification analysis, which reduced the original difference to a nonsignificant trend.

The differences between the two analysts' findings are difficult to interpret and may have been a result of (1) Xie's including site tract-level characteristics in his analytic model or (2) Marcus's having provided especially effective matching of study groups in her model.

➤ On the frequency of placement in *educable mental impairment (EMI) classes* in school, no significant Head Start benefit was found. For both Florida and Colorado, Xie's first-stage unadjusted data analyses found no significant difference for Head Start versus No–Head Start study participants on being placed in educable mental impairment (EMI) classes. Of the Florida Head Start group, 6% had been in EMI classes at some point; this compared with 5% of the No–Head Start group—not a significant difference. At the Colorado site, the proportions of Head Start versus No–Head Start study participants who had ever been in EMI classes were also comparable (5% and 4%, respectively). For both the Florida and the Colorado site, Xie's Model 2 adjusted-data analysis reversed the sign of the Head Start coefficient, indicating a lesser, rather than a greater, likelihood of being in EMI classes. However, the adjusted-data difference for Head Start versus No–Head Start was also not statistically significant. (Marcus elected not to include this variable in her analysis.)

➤ With regard to *high school/GED completion* (treated as simply *high school completion* by both data analysts) or attainment of *some postsecondary education,* no significant Head Start benefit was obtained, but a shift toward a benefit was indicated in the adjusted-data analysis that controlled for differences on the covariates. For the Florida site, after Xie completed the first-stage, unadjusted-data analysis (Model 1), when he employed the set of socioeconomic covariates in a logit regression model (Model 2) to examine high school/GED completion data, he found that the Head Start coefficient shifted from negative (less likely to complete high school/GED) to positive (more likely to complete high school/GED). (See Table 6.1.) However, this indication of a Head Start benefit was not statistically significant. Regarding completion of some postsecondary education, Model 2 also reversed the direction of the Head Start coefficient from negative (less likely to complete some postsecondary education) to positive (more likely to complete some postsecondary education). Similarly, for the two educational variables in Table 6.1, the absolute values of the negative coefficients for Colorado Head Start in Model 1 (first-stage analysis) were also reduced by using the adjusted Model 2 data.

To adjust for initial socioeconomic differences between the Head Start and No–Head Start groups, Marcus employed propensity scores (likelihood of being enrolled in Head Start) in a series of covariance, matching, and subclassification analyses. The first-stage, unadjusted-data analysis found a significantly greater likelihood for the Colorado No–Head Start study participants to have completed high school or the GED and for the Florida and Colorado No–Head Start study participants to have completed some postsecondary education and to have achieved a higher level of education. However, after the analyses had adjusted for initial study-group differences, these findings shifted to being nonsignificant. Also, while the unadjusted-data analysis showed no significant benefit for the Florida Head Start group on completing high school, the adjusted-data analysis improved the odds of a benefit for Head Start, although this Head Start versus No–Head Start difference was nonsignificant.

➤ On the *Adult APL Survey score,* no significant Head Start benefit was found. At both study sites, the first-stage unadjusted-data analyses of both analysts found that in both Florida and Colorado, the No–Head Start group had a significantly higher average Adult APL Survey score than did the Head Start group (see Table 6.1). Upon adjusting the data for the study groups' initial socioeconomic differences (Model 2), Xie did not find significant Head Start versus No–Head Start group differences on Adult APL Survey score at the Florida site. Thus, this statis-

Table 6.1: Head Start Coefficients From Regression Models

Variable	Site	Unadjusted-Data Analysis Model 1	Adjusted-Data Analysis Model 2
High school/GED completion[a]	Florida	-0.31 (0.27)	0.30 (0.38)
	Colorado	-0.88* (0.41)	-0.05 (0.53)
Some postsecondary education[a]	Florida	-0.39 (0.20)	0.05 (0.27)
	Colorado	-1.18* (0.31)	-0.78 (0.42)
Adult APL Survey score[b]	Florida	-2.46* (0.72)	-0.80 (0.78)
	Colorado	-5.19* (1.29)	-3.41* (1.41)
Use of public assistance[a]	Florida	0.45* (0.21)	0.54† (0.29)
	Colorado	0.61* (0.31)	0.38 (0.43)
Arrests[a]	Florida	0.01 (0.22)	-0.21 (0.30)
	Colorado	0.10 (0.30)	-0.45 (0.41)
Convictions[a]	Florida	-0.10 (0.26)	-0.05 (0.32)
	Colorado	0.39 (0.31)	-0.10 (0.42)

Note. Standard errors are presented in parentheses. For *high school graduation/GED, some postsecondary education,* and *Adult APL Survey score,* positive coefficients indicate influence on increasing the levels. For *public assistance, arrests,* and *convictions,* negative coefficients indicate influence on reducing incidence.
[a] Dichotomous outcome variables, logit regression coefficients estimated via maximum likelihood (ML).
[b] Continuous variables, regression slope coefficients estimated via ordinary least squares (OLS).
†$p < .10$. *$p < .05$.

tical model substantially reduced the absolute value of Head Start's negative coefficient and resulted in no significant difference between the study groups on Adult APL Survey score. For Colorado, although the difference between study groups was narrowed by using the adjusted data, the likelihood of a lower Adult APL Survey score for Head Start was still significant.

After employing her adjustment models, Marcus's analysis still found in both Colorado and Florida statistically significant odds of the Head Start group scoring lower than the No-Head Start group on the Adult APL Survey. Again, the analysts proposed that the racial/ethnic imbalance of the Colorado study groups was probably a limiting factor in the analysis. This may especially be the case on this variable, since the Adult APL Survey score may be influenced by language and cultural factors. Yet, on the Florida Adult APL Survey data, Xie found an interesting interaction with race/ethnicity that is noted later in the chapter.

Economic and other life-situation outcomes

➤ **On being *employed/enrolled in school* at the time of the interview, there was some indication of a Head Start benefit at the Florida site.** Marcus's first-stage, unadjusted-data analysis found no significant Head Start benefit on employment/school-enrollment status at the time of the interview. However, after adjusting for the study groups' initial socioeconomic differences, Marcus's analyses found a nonsignificant trend ($p < .10$) toward a Head Start benefit at the Florida site, but no trends or significant differences between study groups at the Colorado site. This variable was not included in Xie's analyses.

➤ **On *teen-parent status,* there was no significant evidence of a Head Start benefit.** Marcus's first-stage, unadjusted-data analysis on the variable teen-parent status found no significant benefit for Head Start. The adjusted-data analyses conducted by Marcus also found no significant effect of Head Start in lessening teen parenthood at either site. Xie did not include this variable in his analyses.

➤ **On *use of public assistance,* no benefit for Head Start was found.** Both analysts' first-stage, unadjusted-data analyses found that at the Florida and Colorado sites, the Head Start groups had a greater likelihood for use of public assistance. As shown earlier, in Table 6.1, Xie's Model 2 analysis controlling for initial socioeconomic differences had some impact on the Florida Head Start coefficient, making it indicate only a nonsignificant trend ($p < .10$) toward more likely use of public assistance. For the Colorado site, Xie's Model 2 unadjusted-data analysis reduced the Head Start group's greater likelihood for using public assistance to being nonsignificant. Marcus's adjusted-data analysis found no statistical difference between the Head Start and No–Head Start study groups on use of public assistance.

➤ **Head Start did not significantly lessen *incidence of arrests* or *incidence of convictions.*** Xie's first-stage, unadjusted-data analyses for the Florida and Colorado sites found no significant benefit of Head Start on the variables of ever having been arrested or convicted, but found a suggested Head Start benefit in Florida on having been convicted. At the Colorado site, on her unadjusted-data analysis, Marcus found a significantly greater number of arrests and convictions for the Head Start group versus the No–Head Start group; she found no study group differences on criminal activity at the Florida site. Xie's Model 2 analysis controlling for the initial socioeconomic differences between groups shifted the Head Start coefficient toward a Head Start benefit in both sites for arrests and in Colorado for convictions, although this was nonsignificant. In her adjusted-data analyses for both sites, Marcus found no significant differences for Head Start versus No–Head Start on arrests or convictions.

Summary of the Head Start Versus No–Head Start Analysis

The findings of the adjusted-data analyses indicate that the Head Start experience may *provide some long-term benefits. These benefits, as assessed in the LTBHS study, are limited, although the direction of the findings observed across the outcome variables and the various analytic models are promising. Xie and Marcus demonstrated analytic models that could be effective in adjusting for initial Head Start versus No–Head Start socioeconomic differences if future quasi-experimental Head Start studies were to utilize a more complete assessment of study groups' initial background differences.*

Analysis of Model Head Start Versus Regular Head Start

This section reports the LTBHS study's findings on core outcome variables for the Florida site's Model (High/Scope) Head Start versus Regular Head Start group. Xie's presentation of initial site tract-level characteristic differences for the Model and Regular Head Start groups in Florida is first examined in this section. (For more detail, see

Table 5.4 on p. 79, Table D.2 on p. 144, and Table D.3 on p. 147.)

➤ **Florida's Model and Regular Head Start groups' long-term outcomes may have been influenced by the two groups' initial socioeconomic differences related to *site tract-level characteristics* and other family-background factors.** Xie's descriptive presentation of the site differences appropriately points to these concerns regarding Model Head Start versus Regular Head Start group comparisons: (1) The urban Model Head Start study participants, as children, lived in areas that were somewhat more advantaged economically than the areas of the urban Regular Head Start study participants. Thus, if the urban Model Head Start subjects were to have better outcomes educationally or otherwise when compared with urban Regular Head Start subjects, it would be unclear if the difference should be attributed to Head Start program variation or to site tract-level characteristic variation. (2) Although, in general, the Florida site's rural areas appeared to be fairly comparable, Florida's rural Regular Head Start area appeared to be somewhat more advantaged than its rural Model Head Start area. Thus, if the rural Regular Head Start subjects were to have better outcomes educationally or otherwise when compared with the rural Model Head Start subjects, it would be unclear if the difference should be attributed to Head Start program variation or to site tract-level characteristic variation. However, as indicated elsewhere, because of the small sample sizes, the rural and urban groups' data had to be combined within the Model and Regular Head Start study groups to compare those latter two study groups on the outcome variables. Thus, the weight of the urban- and rural-area differences may have been lessened, but this could not be directly assessed.

To indirectly assess the potential effects of site tract-level characteristics on Model versus Regular Head Start comparisons, Xie examined all the multiple regression analyses (combining High/Scope and Regular Head Start) on the outcome variables, looking for interactions with the site variables; he found no significant interactions of Head Start and site. His interpretation is that any Model versus Regular Head Start effects cannot be differentiated due to the confound of study group and site characteristics.

It is important to consider that although the variables Xie entered in the analysis indicate some economic characteristics for the U.S. Bureau of the Census tracts, these characteristics do not necessarily pertain to specific households or neighborhoods within each tract. It is therefore difficult to discern the extent to which the site tract-level characteristics data influence outcomes for individuals in the Florida Head Start group within the Florida site's various areas.

As pointed out by Marcus and reviewed in Chapter 5, when we look at family backgrounds of study participants, there were more Model than Regular Head Start study participants with mothers and fathers who had not completed high school (40% vs. 30% completion rate for Model versus Regular mothers; 52% vs. 38% completion rate for Model versus Regular fathers). These differences are not statistically significant, however, and any interpretation of this finding is limited by the fact that these variables are based on retrospective data; we had a considerable amount of missing data on parents' education, especially on father's education. As also noted in Chapter 5, there was a nonsignificant trend toward Model Head Start fathers having completed less education than Regular Head Start fathers. Marcus proposed that since the Regular Head Start group initially appeared to be at no greater risk from the different levels of impoverishment than did the Model Head Start group, better Model Head Start outcomes may reflect actual benefits of the Model Head Start experience.

Thus, in the following sections, with awareness of the apparent variations in site tract-level characteristics, we first present the findings from Xie's Model versus Regular Head Start group analyses, which he conducted for three outcome variables—*GPA*, *arrests*, and *convictions*. We then present findings from Marcus's Model versus

Regular Head Start analyses, including her first-stage, unadjusted-data analyses and her analyses that adjusted for potential Model versus Regular Head Start socioeconomic background differences. Again, few differences are expected between the results of the unadjusted- versus the adjusted-data analyses, since the Model versus Regular Head Start groups' socioeconomic background variables indicated no significant differences and only one nonsignificant trend.

Educational outcomes

➤ **Model Head Start participation, compared with Regular Head Start participation, appeared to lead to a significant likelihood of higher *GPA* from elementary through high school, although *site tract-level characteristics* may have had some influence on this finding.** In their first-stage, unadjusted-data analyses, Marcus and Xie both present data showing a pattern of higher GPA in the elementary, middle, and high school years for the Model (High/Scope) Head Start group, compared with the Regular Head Start group. Looking at GPA by urban and rural sites shows the likelihood of influence of site tract-level characteristic differences. Across Florida's urban and rural site areas, there was a consistent pattern of higher GPA for the Model Head Start groups versus the Regular Head Start groups. In a comparison of just the urban Model Head Start versus the urban Regular Head Start, the Model Head Start advantage appears to lessen in the middle and high school years. In a comparison of just the rural Model Head Start versus the rural Regular Head Start, the advantage remains in favor of rural Model Head Start, even though the rural Model Head Start area showed a greater level of poverty than the rural Regular Head Start area.

Each analyst had a different interpretation of these unadjusted GPA findings. Xie pointed out that there were both site tract-level characteristic differences and parallel urban-rural GPA differences for the No–Head Start study participants. It may also be argued that although the GPA was based on a common grading system applied to the school records data, because the Model and Regular Head Start groups were from different school systems, the criteria for grades, as well as other provisions or experiences, may have differed for Model and Regular Head Start participants. For example, a greater proportion of study participants from the urban Model Head Start area than from the urban Regular Head Start area attended kindergarten. As indicated earlier, Xie concluded that since the analyses could not separate the effects of Model versus Regular Head Start and the effects of urban/rural site-characteristic differences, the results of direct comparisons between the Model and Regular groups in any instance would be very difficult to interpret. Xie therefore elected not to conduct such analyses.

Because there appeared to be some indication that the Model Head Start families initially had lower educational status than the Regular Head Start families, Marcus employed a range of analytic models to adjust for such background differences in her data analyses on GPA. The results of these analyses again revealed significantly higher average GPA for elementary, middle, and high school for Model Head Start versus Regular Head Start (see Table 6.2). Furthermore, the hidden-bias analysis conducted by Marcus indicated that if more unmeasured bias were to be determined and adjusted for, Model Head Start could be found to increase study participants' average GPA by as much as 1 point.

Whereas Xie presented an examination of descriptive data indicating that the socioeconomic characteristics of the Florida site's urban areas may account for the LTBHS study's GPA findings favoring Model Head Start, Marcus pointed out that these LTBHS study findings are consistent with the High/Scope Perry Preschool study's finding of a significantly higher high school GPA for its preschool group versus its no-preschool group (Schweinhart et al., 1993). Furthermore, the LTBHS study's findings of higher GPA for the Model Head Start group parallel the HSPV

Table 6.2: Major Findings for Florida Study Groups

	Model Head Start		Regular Head Start		No–Head Start	
Outcome Variable	**N**	***M*(*SD*)/%**	**N**	***M*(*SD*)/%**	**N**	***M*(*SD*)/%**
Education						
GPA (elementary school)	63	3.15*** (.32)	71	2.38 (.59)	216	2.64 (.56)
GPA (middle school)	65	2.53*** (.66)	70	2.00 (.72)	220	2.26 (.63)
GPA (high school)	60	2.31*** (.82)	69	1.75 (.77)	223	1.98 (.76)
High school/GED completion	74	78%	86	85%	264	86%
Some postsecondary education	74	32%	86	47%+	264	50%
Highest year of education completed	74	11.92 (1.77)	86	12.24 (1.51)	264	12.50 (1.32)
Adult APL Survey score	69	25.70 (7.78)	80	27.21 (7.47)	246	29.0 (6.45)
Other life situations						
Employed/in school at time of interview	74	70%	86	74%	264	68%
Teen parenthood	74	26%	83	28%	261	20%
Ever on public assistance	74	46%	86	41%	264	33%
Arrests	74	.95+ (2.30)	86	1.72 (3.46)	264	1.25 (3.46)
Convictions	74	.54* (1.59)	86	1.41 (3.40)	264	.95 (3.19)

Note. Statistical comparisons shown here are conducted for Model versus Regular Head Start; the No–Head Start group data are provided only for descriptive purposes. *M* = mean; *SD* = standard deviation. Group differences were tested on the data that were unadjusted for initial family socioeconomic variables, using *t*-test and χ^2, as appropriate; the same pattern of findings and levels of significance reported here were found in subsequent analyses with the data adjusted for the initial socioeconomic differences.

+$p < .10$. *$p < .05$. ***$p < .001$.

study's finding of gains on intelligence test scores lasting into elementary school for its Model (High/Scope) Head Start participants.

➤ **On other educational outcomes—*high school/GED completion* and *highest year of education* completed—no significant differences between the Model Head Start and the Regular Head Start groups were found. There was a nonsignificant trend for the Regular Head Start group to be more likely to attain *some postsecondary education*.** The first-stage, unadjusted-data analysis employed by Marcus found no significant differences between Model and Regular Head Start on completion of high school/GED (78% vs. 85%, respectively) and highest year of education completed (*M* = 11.9 years [*SD* = 1.8] vs. *M* = 12.2 years [*SD* = 1.5], respectively). The adjusted-data analysis also found no significant differences for these comparisons. Furthermore, Marcus found that a moderately large imbalance between Model and Regular Head Start groups due to an unmeasured covariate would be necessary to shift the direction of the group findings.

Marcus did find the nonsignificant trend ($p < .10$) of a greater percentage attaining some postsecondary education for the Regular versus Model Head Start group (47% vs. 32%) on two out of three of her adjusted-data analyses. On the other hand, Marcus's hidden-bias analysis indicated that the Model Head Start group's level of education could possibly increase by as much as ½ year if a moderately large unmeasured covariate were found to be masking a Model Head Start effect. For example, site-related socioeconomic differences or other factors (such as differences in school systems, in rates of kindergarten participation, or in ages of study participants at interview) in one of the urban Model or urban Regular Head Start areas may have influenced the obtained trend. It may be that family-background differences have an especially important effect on a young person's pursuit of postsecondary education.

➤ **On *Adult APL Survey score,* two out of the three Model versus Regular Head Start analyses revealed no significant differences; one analysis did find a significant difference indicating a higher average *Adult APL Survey score* for the Regular Head Start group.** Marcus's first-stage, unadjusted-data analysis and her adjusted-data analyses were generally in agreement that there was no significant difference between Model and Head Start study participants on Adult APL Survey score. One of her adjusted-data analyses, a subclassification analysis, controlled for systematic socioeconomic background differences between the Model and Regular Head Start groups by determining subclasses of study participants within each group who were comparable on the covariates. This subclassification analysis indicated that the Regular Head Start group had a higher average Adult APL Survey score. However, Marcus pointed out that given the pattern of no significant differences on the adjusted-data analyses she conducted, the finding may have been due to the distribution properties of the Adult APL Survey scores and the subclassification structure.

Other life-situation outcomes

➤ **On several life-situation outcomes—use of *public assistance, teen-parent status,* and *employed/enrolled in school* at time of the follow-up interview—no significant differences were found for the Model Head Start group versus the Regular Head Start group.** Marcus's first-stage, unadjusted-data analysis indicated no significant differences in the proportions of Model versus Regular Head Start participants who had ever received public assistance (46% vs. 41%, respectively) or who had been a teen parent (26% vs. 28%, respectively). There was also no significant difference between these two Head Start subgroups in their proportions employed/enrolled in school at the time of the follow-up interview (70% vs. 74%, respectively); the adjusted-data analyses also found no significant differences.

➤ **The Model Head Start group, compared with the Regular Head Start group, had significantly fewer *convictions* and a nonsignificant trend of fewer *arrests,* although site tract-level characteristics may have had an influence on these findings.** Xie's descriptive examination of the unadjusted data on study participants' incidence of arrests and convictions, aggregated across the urban Florida areas, revealed a pattern of a smaller proportion arrested and a smaller proportion convicted from the Model Head Start group versus the Regular Head Start group. This was especially the case for the urban areas; Florida's rural areas, however, had a pattern of a smaller proportion arrested and a smaller proportion convicted from the Regular Head Start group. These different patterns suggest that the incidence of arrests and convictions may have been influenced by socioeconomic site characteristics or other site factors, such as differences in school-system resources or school-system opportunities (such as the opportunity for kindergarten).

Also to be considered was the potential influence of study participants' being different ages across the urban and rural areas at the time of the follow-up interview. However, the study's use of a common time frame in Florida for the collection of court records data—1 to 2 years after the interview phase had ended—considerably limited this potential influence. Also, the criminal activity composite was constructed largely from the court records data, with only a small number of additions from the self-report interview data. (See Chapter 3, p. 45, and Chapter 4, p. 64, for further details.)

Whereas Xie focused on the *incidence* of arrests and convictions (whether or not a participant was ever arrested or convicted), Marcus used the *number of arrests* and *number of convictions* as variables in her statistical analyses of the criminal-activity outcome variables. Marcus's first-stage, unadjusted-data analysis and her subsequent matching (adjusted-data) analysis indicated a nonsignificant trend of fewer arrests for the Model Head Start group versus the Regular Head Start group (see Table 6.2). On her two other adjusted-data analyses—the covariance-adjusted and the subclassification analyses—a nonsignificant trend of fewer arrests for the Model Head Start group ($p < .10$) was also found.

Marcus found significantly fewer criminal convictions for the Model Head Start group versus the Regular Head Start group in her first-stage, unadjusted-data analysis and across all her adjusted-data analyses (see Table 6.2). Although there may have been additional influences, as just noted, this finding of a lessening of criminal convictions in young adulthood is consistent with other studies that have examined the long-term effects of high-quality preschool programs (e.g., Lally et al., 1988; Schweinhart et al., 1993; Schweinhart & Weikart, 1997).

Summary of the Model Head Start Versus Regular Head Start Analysis

The comparison of the Model Head Start versus Regular Head Start groups sought to determine if there was any advantage to having participated in a child development, theory-based program (as opposed to a "regular" program). In this case, the Model Head Start program used the High/Scope preschool curriculum. Although there were site-characteristic differences to consider—differences inherited from the Head Start Planned Variation study design—the pattern of higher GPA, *especially in the elementary school years, and fewer* arrests *and* convictions *for the Model Head Start group is both intriguing and consistent with findings from other long-term studies of the High/Scope curriculum. The absence of any other pattern of significant findings for the Model versus Regular Head Start groups may indicate the influence of both site tract-level characteristic differences and other, unmeasured family-background differences. Yet, even with such influences, the analyses gave some indication of potential long-term benefits to be gained from participation in the Model Head Start program.*

Comparing Other Specific Subgroups—Analysis by Gender

As discussed, after his first-stage, unadjusted-data analysis (Model 1), Xie employed his second-stage analyses (Models 2 and 3). While his Model 2 controlled for Head Start and No–Head Start study participants' initial socioeconomic differences, his Model 3 also examined the extent to which the main study group variable (Head Start versus No–Head Start membership) interacted with other group memberships—*gender, race/ethnicity,* and *socioeconomic level.* Marcus also examined the potential effects of Head Start for these subgroups in her adjusted-data analyses. This section reports on the analysis by *gender.*

In the following analyses of Head Start versus No–Head Start groups by *gender,* for the most part, we focus on data analyses that adjusted for the study groups' initial socioeconomic differences, although we sometimes give unadjusted-data analysis findings to illustrate the shift in findings when the covariates are included in the analyses. In some analyses, some coefficients that are found to be negative (including significant and nonsignificant effects) should be taken not as indicative of harmful outcomes but as indicative of less likelihood of providing a positive influence on an outcome, as Xie argued in the case of a negative coefficient found for males (see Xie, Appendix D, p. 152, for a discussion of Manski, 1995). Further, Xie proposed that any obtained negative effects in the analyses are likely due to residual socioeconomic differences between the Head Start and No–Head Start study groups not accounted for in Xie's Model 3 or in Marcus's adjusted-data analysis. Marcus examined the potential influence of various combinations of *gender* and *race\ethnicity* (e.g., African American males and females compared with White/other males and females) for a range of socioeconomic levels. Marcus's examination of Head Start versus No–Head Start *gender* differences was primarily conducted to establish a basis for directing future studies, whereas Xie's analyses were the more extensive ones. Thus, discussion of the findings on *gender* is based mostly on Xie's analysis.

Educational outcomes

➤ **On *high school GPA,* Head Start provided a nonsignificant trend toward a benefit for Florida females. No Head Start benefits were found on *GPA* at any level for males in Florida or for either males or females in Colorado.** Although overall there was no significant Head Start benefit on elementary, middle, or high school GPA in Colorado or in Florida, Xie's adjusted-data analyses showed Florida Head Start females, as compared with Florida No–Head Start females, to have a nonsignificant trend of higher GPA in high school (see Table 6.3). No benefit on GPA was found for Head Start males in Florida or for either Head Start males or Head Start females in Colorado.

➤ **On *high school/GED* completion, Head Start provided a significant benefit for Florida females. No Head Start benefit on this variable was found for males in Florida or for either males or females in Colorado.** Although overall there were no significant Head Start effects on high school/GED completion, Xie's adjusted-data analysis found that significantly more Florida Head Start females, as compared with Florida No–Head Start females, completed high school/GED. As shown in Table 6.4, these results parallel findings from the High/Scope Perry Preschool study on high school/GED completion.

At the Colorado site, Xie found no significant differences for Head Start versus No–Head Start females on high school/GED completion. Both Xie and Marcus proposed that this could be due to the extent of the racial/ethnic imbalance and related socioeconomic differences found for the Head Start versus No–Head Start groups in Colorado, as discussed in Chapter 5, p. 71, and presented in Table 5.2.

➤ **On attainment of *some postsecondary education,* Head Start produced a nonsignificant trend toward a benefit for Florida females. No indication of a Head Start benefit on this variable was found for males in Florida or for either males or females in Colorado.** Although overall there were no significant benefits for Head Start on attainment of some postsecondary education, there was a nonsignificant trend ($p < .10$) indicating that more Head Start than No–Head Start females in Florida had completed some postsecondary education. There were no significant benefits on this variable for Florida Head Start males or for either Colorado Head Start males or Colorado Head Start females. See Table 6.4, which illustrates how these findings compare with key results from the High/Scope Perry Preschool study.

➤ **On *Adult APL Survey score* or placement in *EMI classes,* no significant Head Start benefit was found for females or for males in either Florida or Colorado.** No significant differences were found in Xie's Model 3 analysis for the interaction of gender and Head Start versus No–Head Start status on the variables EMI class placement or Adult APL Survey score.

Other life-situation outcomes

➤ **On incidence of *public assistance,* there was no significant Head Start benefit for either females or males in either Florida or Colorado.**

Table 6.3: Florida Head Start Versus No–Head Start Females—Analysis Using Unadjusted and Adjusted Data

	Unadjusted-Data Analysis		Adjusted-Data Analysis	
	Head Start Females	**No–Head Start Females**	**Head Start Females**	**No–Head Start Females**
High school/GED completion (%)	84.8	82.7	95.1*	81.1
Some postsecondary education (%)	46.8	51.9	64.3'	51.7
GPA (elementary school) (*M*)	2.92	2.76	2.77	2.82
GPA (middle school) (*M*)	2.44	2.34	2.46	2.40
GPA (high school) (*M*)	1.98	2.20	2.18'	2.05
Use of public assistance (%)	59.5	48.9	66.0	56.3
Ever arrested (%)	10.1	18.0	5.0*	14.9
Ever convicted (%)	7.6	12.0	6.4	9.6

Note. The means (*M*) and percentages of the adjusted-data analysis are derived from Xie's Model 3 series of multiple regression analyses, summarized in Table D.14 in Appendix D. The unadjusted-data analysis figures are presented for descriptive purposes. For the adjusted-data analysis, the significance levels pertain to group by gender interactions.
'$p < .10$. *$p < .05$.

Table 6.4: Females' Completion of High School/GED and Some Postsecondary Education in the LTBHS and High/Scope Perry Preschool Studies

Variable/Study	Program Group Females	Comparison Group Females
High school GED completion (%)		
LTBHS Study—Florida	95.1*	81.1
High/Scope Perry Preschool Study	84.0***	35.0
Some Postsecondary Education(%)		
LTBHS Study—Florida	64.3'	51.7
High/Scope Perry Preschool Study	44.0**	27.0

Note. LTBHS = Long-Term Benefits of Head Start. The LTBHS study data presented here were adjusted for the study groups' initial socioeconomic differences. Statistical comparisons shown here were conducted between the study groups *within* each study.
'$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

The unadjusted- and adjusted-data analyses provided no evidence that Head Start, overall, lessened the incidence of public assistance. Also, as shown in Table 6.3, Xie's Model 3 analysis of Florida data found no Head Start benefit for females on this variable. Likewise, his Model 3 analysis showed no significant Head Start benefit on this variable for Florida males, for Colorado males, or for Colorado females.

➤ **On incidence of *arrests,* there was a Head Start benefit for females in Florida; compared with No–Head Start females, they had a**

significantly lower incidence of arrests. There were no significant differences on *convictions* for Head Start versus No–Head Start females. There were no significant differences on *arrests* or *convictions* for Head Start versus No–Head Start Males in Florida or for Head Start versus No–Head Start males or females in Colorado. In his overall adjusted-data analyses for the Florida and Colorado sites, Xie found that the regression coefficients were in the direction of Head Start lessening the incidence of criminal arrests and convictions for males and females, but this finding was not statistically significant. However, a significant difference was obtained at the Florida site on the adjusted-data analyses: Compared with No–Head Start females, significantly fewer Head Start females were found to have arrests (see Table 6.3). Incidence of criminal convictions was somewhat lessened for Head Start females, compared with No–Head Start females, in Florida and Colorado, but these differences were not significant.

Summary of the Analysis by Gender

The LTBHS study found higher educational achievement and less criminal activity for Florida Head Start females. The higher educational achievement included higher GPA in high school, *a greater percentage completing* high school/GED, *and a greater percentage attaining* some postsecondary education. *These findings parallel some of the findings of the High/Scope Perry Preschool study. Although such findings provide evidence that Head Start has benefits, further consideration should be given to why this positive impact appears to be more likely for females—and only for females in one site.*

Comparing Other Specific Subgroups—Analysis by Racial/Ethnic Membership and Socioeconomic Level

Both Xie and Marcus examined the extent to which a study participant's membership in various other subgroups may affect Head Start's potential benefits. Putting aside whether or not a study participant was in Head Start, their data analyses also examined the extent to which being of a particular racial/ethnic group or of a particular socioeconomic level affected the study participant's educational outcomes and other life-situation outcomes. In the following section, we discuss the patterns of findings on outcome variables from interactions of some of the various subgroups with the Head Start versus No–Head Start comparison.

Educational outcomes

➤ On the *elementary school GPA, high school/GED completion,* and *highest year of education* completed, there were indications that Head Start benefits were related to *socioeconomic level* or *race/ethnicity*. At the Florida site, Xie found no significant difference between African Americans and Whites/others on the likelihood of high school/GED completion. However, at the Colorado site, the likelihood of high school/GED completion was significantly lower for Hispanic Americans than for Whites/others. Xie's analyses indicated that mother's and father's higher levels of education greatly increased a study participant's odds of high school/GED completion and attainment of some postsecondary education in Florida, but parents' education levels had no significant effect in Colorado. Greater family use of welfare significantly decreased a study participant's odds of high school/GED completion and attainment of some postsecondary education in Florida but had no significant effect in Colorado.

To examine subgroup interactions with respect to the Head Start versus No–Head Start groups on the core educational outcomes, Marcus examined odds ratios by propensity score quintiles (see Table F.10 on p. 196 and Table F.11 on p. 197; note that lower quintiles reflect lower socioeconomic status). She found a pattern suggesting that Colorado study participants with the lowest economic levels benefited the most from Head Start on their elementary school GPA, high school/GED completion, and highest year of education completed. At the Florida site, Marcus's analysis indicated that Head Start was significantly more beneficial for the African American females than for the females from the White/other group in obtaining higher GPA in elementary school. The Head Start White/other group in Florida tended to have lower GPA in elementary and middle school, when compared with the No–Head Start White/other group. Because these subgroups of the main study groups involved small sample sizes, these results are only suggestive.

➤ **On *Adult APL Survey score,* Florida Head Start appeared to provide a benefit related to *racial/ethnic membership*.** Xie found a sizable and nonsignificant trend for the interaction between Head Start and racial/ethnic membership at the Florida site. This interaction finding suggests that participation in Head Start may improve the Adult APL Survey score for African Americans by several points.

Other life-situation outcomes

➤ **On incidence of *public assistance,* there appeared to be a Head Start benefit in Colorado for those study participants from the most economically disadvantaged backgrounds.** On the life-situation outcomes, to examine subgroup interactions with respect to Head Start versus No–Head Start, Marcus examined odds ratios by propensity score quintiles (see Table F.10 on p. 196; note that lower quintiles reflect lower socioeconomic status). She found that at the Colorado site only, Head Start study participants from the most economically disadvantaged backgrounds were the least likely to have used public assistance in young adulthood. In Florida, Whites/others in Head Start versus No–Head Start were significantly more likely to have used public assistance as young adults. Marcus speculated that this finding may reflect the different processes by which Whites/others versus African Americans had been recruited into Head Start. It is difficult to interpret such findings, however, because the size of the White/other subgroup was not large in this study. Xie's Model 3 analysis found no interaction effects on the incidence of public assistance for racial/ethnic membership or for levels of socioeconomic background.

Summary of the Analysis by Racial/Ethnic Membership and Socioeconomic Level

The odds of Head Start providing educational benefits may be stronger for some racial/ethnic groups and for the lowest socioeconomic groups. Also, the odds of Head Start enhancing the life situations of young adults may be stronger for some racial/ethnic groups and for the lowest socioeconomic groups.

The analyses for interaction of racial/ethnic membership *and* socioeconomic level *with Head Start versus No–Head Start comparisons indicated that Head Start appeared to have provided some benefits to African Americans at the Florida site. The imbalance of Hispanic Americans in the Head Start and No–Head Start groups at the Colorado site limited our ability to estimate potential benefit with this racial/ethnic group. Head Start benefits for Whites/Others were difficult to evaluate, partly because of their low numbers in the Head Start groups in this study. The fact that Head Start appears to provide some benefit to those who initially had the lowest socioeconomic status suggests that Head Start appears to be serving those most in need.*

CONCLUSION

Although Xie and Marcus's separate, independent analyses yielded convergent as well as some divergent results, the overall picture is encouraging: If the design of a Head Start long-term study can effectively control for study groups' initial background differences, there is some reason to expect the study to find long-term benefits of Head Start participation. We may also expect to find some benefits resulting from Head Start programs using such models as the High/Scope preschool curriculum and other child development approaches.

The Colorado site, where the Head Start group contained a significantly greater proportion of Hispanic Americans than did the No–Head Start *group, demonstrated that Head Start long-term benefits were plausible, but the imbalance in the study's design severely limited the potential of our analysis methods to adjust for the study groups' background differences. In urban and rural areas like those we studied in Florida—where the Head Start and No–Head Start groups had the most initial similarity—the most likely long-term benefits would appear to involve the variables* high school/GED completion, *attainment of* some postsecondary education, high school grade point average, *and* incidence of arrests *and* convictions.

7

SUMMARY AND IMPLICATIONS

The Long-Term Benefits of Head Start (LTBHS) study, conducted at two sites—one in Florida and one in Colorado, compared two groups of young adults at age 22. At each site, one study group—the Head Start group—consisted of former Head Start children, most of whom had participated in the 1970 Head Start Planned Variation study; the other study group—the No–Head Start group—consisted of young people who as children had not participated in Head Start or any other preschool program. In both Florida and Colorado, the study participants from both study groups had spent their early childhood years in the same communities.

When interviewed at age 22, the LTBHS study's 622 participants were in a variety of circumstances—attending community colleges or 4-year colleges or universities, working and raising families, looking for work while on welfare assistance, serving prison sentences, or serving in the military. The study interviewers collected extensive information on the study participants' educational experiences and their other life circumstances. With subjects' permission, information was also collected from their school records—kindergarten through postsecondary—and from public assistance and criminal records, as applicable. This final chapter recaps the LTBHS study and considers its implications for today's Head Start program development and research.

WHAT STUDIES HAVE THUS FAR REVEALED ABOUT HEAD START EFFECTS

Since Head Start began in 1965, educators, researchers, and policymakers have asked, *Does Head Start work?* One early study of Head Start's initial program (conducted by the Westinghouse Learning Corporation and Ohio University, 1969) found the program to have few effects, and these findings have continued to haunt Head Start research over the past 30 years—despite the critiques pointing out the study's limitations and despite the more positive results obtained in reanalyses of the study's data.

However, since 1969 numerous other studies have found Head Start to produce a variety of *short-term* positive effects—enhanced intelligence and achievement test scores, reduced grade-retention, and better school-adjustment (see McKey et al., 1985 for a summary). These accumulated studies gradually led to a growing acceptance of the idea that Head Start produces early benefits. So, many educators and policymakers then began to ask a new question: *Do the effects of Head Start last?*

A number of studies that followed former Head Start children into their later elementary school years found what has been termed a fade-out of the early effects of Head Start (see Barnett, 1992). Yet various investigators and methodologists (Barnett, 1992; Campbell & Frey, 1970; Powell & Roane, 1996) have pointed out that unless comparable types and levels of educational experience are provided following the Head Start experience—throughout children's school years—Head Start benefits are not likely to persist, or certainly not likely to persist to the same extent. Thus, findings showing that early Head Start effects diminish over time should not be taken to invalidate the findings of early effects.

As suggestive of Head Start's potential to have long-term effects on the lives of children from low-income backgrounds, many have cited the positive long-term findings of two High/Scope longitudinal studies—the High/Scope Perry Preschool study and the High/Scope Preschool Curriculum Comparison study (Schweinhart et al., 1993; Schweinhart & Weikart, 1997). Although the High/Scope preschool approach was one of the Head Start Planned Variation study's curriculum models in the 1970s and has been adopted as a curriculum approach in many Head Start programs from that time up to the present, these well-known High/Scope studies were conducted on preschool demonstration programs, not on actual Head Start programs.

Newspaper editorials have sometimes cited a glaring lack of evidence of Head Start's long-term effects. This criticism was echoed in a U.S. General Accounting Office report (1997). Although Head Start studies have documented effects of Head Start occurring up to 3 or 4 years after the program, as Edward Zigler noted in *The New York Times* (July 24, 1993), "It is hard for proponents to argue convincingly for more dollars because there are no studies showing how productive Head Start children have been as adults: The appropriate research has not been done." To date, in fact, very few early childhood programs of *any* kind have been subjected to long-term assessment. In addition to the two High/Scope studies just mentioned, a few other studies have followed up participants in high-quality early childhood programs and have determined that over the long run, children from the lowest socioeconomic levels can achieve a higher level of education than they would have without their preschool experience (Campbell & Ramey, 1994; Consortium for Longitudinal Studies, 1983; Lally et al.,1988).

At Head Start's inception, many educators, child development specialists, and policymakers looked to the program as a potential inoculation against the disadvantages of poverty. Then, in light of the research showing that Head Start's early effects tended to diminish over time, this expectation began to look somewhat optimistic. Now, with the aim of achieving longer-lasting effects such as those found in the High/Scope Perry Preschool study, Head Start program development and research is increasingly focused on enhancing the *quality* of Head Start programs—in terms of curriculum, staff development, parent involvement, and relationships with schools and communities. In this context, the LTBHS study provides new insights for today's researchers by providing a better understanding of how Head Start may affect the long-term progress of young people born into poverty. The study also provides important methodological directives for future long-term studies of Head Start.

AN OVERVIEW OF THIS STUDY

The Methodology

The main research question of the LTBHS study is this: *What differences does Head Start make in the lives of its participants?* Ideally, the best estimate of how former Head Start study participants would have fared without Head Start would be accomplished by comparing two randomly assigned groups, one of which subsequently attended Head Start and one of which did not, as was done in a few experimental studies of preschool versus no-preschool children (e.g., the Abecedarian Project, Campbell & Ramey, 1994, 1995; the High/Scope Perry Preschool study, Schweinhart et al., 1993). Because the Head Start groups (in Florida and Colorado) in the LTBHS study were drawn from the earlier, quasi-experimental Head Start Planned Variation (HSPV) study (Smith, 1973), the LTBHS study design was necessarily quasi-experimental (i.e., it also employed nonrandom program and comparison groups). Thus, at each site of the LTBHS study, as indicated previously, the Head Start group was compared with a No–Head Start group in order to assess the overall impact of Head Start. Also, to assess the potential advantage of using a child development, theory-based approach in the Florida site's Head Start program, at that site a Model Head Start group (which employed the High/Scope curriculum) was compared with a Regular Head Start group.

Because the Head Start program staff at both study sites had recruited children from the communities' poorest families (as they are encouraged to do), the No–Head Start study participants were somewhat better-off economically, even though they came from the same communities as the Head Start children. Overall, the LTBHS study had a number of methodological limitations:

- Few valid socioeconomic status measures at study entry
- Retrospective reporting of the study groups' initial socioeconomic status
- Initial differences between the Head Start and No–Head Start groups in site tract-level characteristics
- Initial differences between the Head Start and No–Head Start groups in family socioeconomic characteristics
- Different proportions of Hispanic Americans and Whites/Others in the Head Start and No–Head Start groups at the Colorado site
- Likely additional hidden sources of bias in subjects' selection into the Head Start or No–Head Start groups
- Likely differences between the Model and Regular Head Start groups in site tract-level characteristics

At each of the LTBHS study's two sites, the LTBHS study groups were these:

- **The Head Start Group**—a group of young people who had participated in Head Start programs in 1970–71; at the Florida site these Head Start participants consisted of a subgroup who had participated in a Head Start program based on child development theory (the High/Scope preschool model) and another subgroup who had participated in a Head Start program using an approach with no particular theoretical base. These two subgroups are called, respectively, the *Model Head Start group* and the *Regular Head Start group*.
- **The No–Head Start Group**—a comparison group of young people who had never participated in Head Start or in any other early childhood education program.

Of the LTBHS study's pool of identified potential study participants, approximately 77% were interviewed at 22 years of age (about 70% of the Colorado site pool and about 80% of the Florida site pool), with comparable proportions of study participants across study sites and groups. In its final sample, the LTBHS study had 622 study participants—337 African Americans, 117 Hispanic Americans, and 168 Whites/others.

An advisory panel of national experts in methodology and statistics made recommendations for the selection of core variables and statistical approaches for the data analyses. To adjust for the study groups' original socioeconomic differences, two independent data analyst/scholars—Sue Marcus of the University of Pennsylvania and Yu Xie of the University of Michigan—employed a range of statistical models. Thus, although it did not employ random design, the LTBHS study was able to estimate how the Head Start group of study participants would have fared without Head Start.

The Expected Findings

The High/Scope Perry Preschool study, which employed the High/Scope preschool model with children from low-income families (see Schweinhart et al., 1993), found long-term benefits for its preschool group versus its no-preschool

BEST COPY AVAILABLE

group. We expected to find a similar direction or pattern of effects for the Head Start groups in the LTBHS study, since the core study variables we selected for analysis (pertaining to educational achievement, economic independence, and incidence of criminal activity) were many of the same ones employed in the High/Scope Perry Preschool study. Because that study had used random assignment to study groups, only one significant socioeconomic-background differences was found in comparing its preschool and no-preschool groups at study entry; the virtual equivalence of these groups at study entry contributed to obtaining a good estimate of the effects of preschool participation. By contrast, considering the significant socioeconomic-background differences between the LTBHS study's Head Start group and No–Head Start group at study entry, we expected that a good estimate of the effects of Head Start would be more difficult to achieve. However, to the extent that the statistical models employed by the data analysts could effectively adjust for the Head Start versus No–Head Start groups' initial differences, we expected to find some indications of long-term benefits for Head Start participation. Because *full* adjustment for the Head Start group's initial measured socioeconomic disadvantage (and for the hidden study biases inherent in the nonrandom design) was not likely, we expected to find limited long-term benefits of Head Start.

Whatever the magnitude of the LTBHS study's obtained effects, we expected to find a direction or pattern of results indicating greater educational achievement and improved life circumstances for the Head Start Group, as was found for the program groups in the few other studies of economically disadvantaged children who participated in model early childhood programs, (e.g., Campbell & Ramey, 1995; Lally et al., 1988; Reynolds, Chang, & Temple, 1997; Schweinhart et al., 1993; Schweinhart & Weikart, 1997). For analysis, based on the High/Scope Perry Preschool study's positive findings, we selected from the LTBHS study database a set of variables most likely to find positive effects for Head Start, for Model Head Start, and for Head Start females. The specific expected findings for Head Start study participants were *higher* GPA, incidence of high school graduation, level of education, APL Adult Survey score, incidence of employment or enrollment in school, and *less* placement in educable mental impairment (EMI) classes, incidence of public assistance, incidence of being a teen parent, and incidence of criminal arrests and convictions.

Further, based on the High/Scope Preschool Curriculum Comparison study's positive outcomes for its High/Scope program and its traditional Nursery School program (Schweinhart & Weikart, 1997), we also expected to find positive outcomes for the LTBHS study's Model (High/Scope) Head Start group. Because the High/Scope studies and some others found better educational outcomes for females with preschool participation, the LTBHS study's analyses by gender were expected to reveal stronger educational benefits for females who had participated in Head Start.

A SUMMARY OF THE LTBHS STUDY'S MAJOR FINDINGS

The following summary of the LTBHS study's major findings represents convergent findings across the several statistical models employed by analysts Xie and Marcus (see Appendixes D, E, and F for their reports).

Both analysts initially identified the LTBHS study's methodological limitations, which they judged would result in *a likely underestimation of Head Start benefits*. The data analysts' statistical models estimated the extent of the No–Head Start study participants' original socioeconomic advantage over the Head Start participants. *Both data analysts presented differences between the Head Start and No–Head Start groups' initial socioeconomic backgrounds, thus demonstrating that it would be invalid to treat the unadjusted data as the source for estimating the effects of Head Start.* The data analysts subsequently adjusted the data to compensate for the original socioeconomic differences and thereby achieve a fairer comparison for the Head Start group.

Estimates of Head Start's Long Term Effects

Although most often the LTBHS study findings were not of the expected magnitude or statistical significance, *the pattern of findings across most of the educational and life-situation outcome variables (especially at the Florida site) has the same direction as the pattern of significant findings obtained in the High/Scope Perry Preschool study and in the High/Scope Preschool Curriculum Comparison study.* At the LTBHS study's Colorado site, the Head Start versus No–Head Start findings were often in the expected direction; however, compared with the Florida site, the Colorado site showed a lack of significant Head Start benefits. Most likely this was due to the greater initial socioeconomic differences between the Head Start and No–Head Start groups in Colorado. Also, a majority in Colorado's Head Start group were Hispanic American (and a majority in its No–Head Start group were White/other), and the employed statistical models perhaps could not compensate for any additional disadvantages that the Colorado Head Start group may have had to surmount because of this imbalance (e.g., linguistic obstacles, hiring descrimination).

The convergence of Xie's and Marcus's patterns of findings indicates that if the LTBHS study design or analyses of data had been able to control more effectively for initial group differences, significant Head Start long-term benefits—especially higher educational attainment and lower incidence of criminal activity—would likely have been found. Using data adjusted for the study groups' socioeconomic differences, both analysts obtained findings that pointed toward such potential Head Start benefits, but these findings were generally not statistically significant. One exception to the pattern was a significant finding of a higher average *Adult APL Survey score* for the Colorado No–Head Start group, compared with the Head Start group. The analysts also noted two nonsignificant trends ($p <$ .10) at the Florida site—a suggested No–Head Start advantage (less likelihood of using *public assistance*) and a suggested Head Start advantage (a greater likelihood of being *employed/enrolled in school* at the time of interview).

Findings for Specific Subgroups

Despite the LTBHS study's methodological obstacles, some analyses of outcomes for the study subgroups revealed specific significant findings similar to findings of the High/Scope Perry Preschool study. For example, the *adjusted-data analyses* by gender revealed significant differences for Head Start females versus No–Head Start females at the Florida site:

- A higher proportion of Head Start females completed *high school/GED* (95% vs. 81%, $p < .05$).
- A smaller proportion of Head Start females had ever been *arrested* (5% vs. 15%, $p < .05$).

Also in Florida, two nonsignificant trends ($p < .10$) noted in these analyses suggested benefits for Head Start females—they had a higher proportion achieving *some postsecondary education* and a higher *high school GPA*, when compared with No–Head Start females; another nonsignificant trend suggested higher *Adult APL Survey scores* for Head Start African Americans versus No–Head Start African Americans.

Also at the Florida site, both the adjusted- and the unadjusted-data analyses revealed these significant differences for the Model (High/Scope) Head Start group versus the Regular Head Start group:

- The Model Head Start group had a significantly higher ($p < .001$) *elementary school GPA* ($M = 3.15$ [$SD = .32$] vs. $M = 2.38$ [$SD = .59$]), *middle school GPA* ($M = 2.53$ [$SD = .66$] vs. $M = 2.00$ [$SD = .72$]), and *high school GPA* ($M = 2.31$ [$SD = .82$] vs. $M = 1.75$ [$SD = .77$]).

- The Model Head Start group had significantly fewer ($p < .05$) criminal *convictions* ($M = .54$ [$SD = 1.59$] vs. $M = 1.41$ [$SD = 3.40$]).

Two nonsignificant trends ($p < .10$) were also noted in these adjusted- and unadjusted-data analyses: One suggested a benefit for Regular versus Model Head Start, with a greater proportion having achieved *some postsecondary education*, and one suggested a benefit for the Model Head Start group, with a lower number of criminal *arrests*.

Assessing the Findings

Although we had expected to find significantly higher educational levels and significantly better life-situations for the overall group of Head Start study participants, the pattern of LTBHS study findings for subgroups of the Head Start group was in the direction of higher educational outcomes and less criminal activity, and this is similar to the direction of High/Scope Perry Preschool study findings. Although differences in site tract-level characteristics (e.g., the differences in poverty rates) inherited from the original HSPV study design had to be considered in the LTBHS study's analyses, the positive findings that did result in the LTBHS study suggest that using the High/Scope curriculum or another child development, theory-based curriculum in a Head Start program may be especially likely to produce long-term benefits.

The LTBHS study found no significant Head Start benefit regarding reduction of placement in *EMI classes*, use of *public assistance*, or *teen-parent status* and no significant Head Start benefit regarding improvement of *Adult APL Survey score*. Both Xie and Marcus suggested that problems in measuring these variables may have prevented achieving a true estimate of effects. Some possible explanations for this lack of significant findings are these: It is difficult to speculate about frequency of placement in *EMI classes*, in part because this variable would seem to be especially affected by differences in school-system policies and facilities. With respect to *Adult APL Survey score*, some interviewers reported that administering the *Adult APL Survey* at the time of the follow-up interview was difficult—some study participants did not really want to complete the survey. Also, in Colorado, where the Head Start group appeared to have a lower average *Adult APL Survey score* than did No–Head Start group, there may have been a language-comprehension disadvantage for the Hispanic Americans (who were the majority of the Head Start group but the minority of the No–Head Start group). In both Florida and Colorado, for the variables *public assistance* and *teen-parent status* in particular, making the transition to young adulthood may have also been especially affected by the economic circumstances of the family, which were found to be somewhat different for Head Start versus No-Head Start study participants.

Although our various speculations about these variables may have some validity, Head Start's long-term effects may simply not extend to these variables, especially the ones having to do with life-situation outcomes. Furthermore, if Head Start does not have a strong, lasting impact on its participants' educational achievement, its impact on their life situations as adults may be even more limited.

In addition to Colorado's racial/ethnic imbalance between study groups, there is another possible explanation for the lack of significant findings at that site, and this explanation has its origins in the Head Start Planned Variation study (Lukas & Wohlleb, 1972; Smith, 1973). In the HSPV study, compared with the Florida Model Head Start staff, the Colorado Model Head Start staff had less preservice training, and in the HSPV findings, their students subsequently showed fewer gains in intelligence test scores over the Head Start year. (Unfortunately, as explained in an earlier chapter, at the time of the LTBHS study, we were not able to determine Model versus Regular Head Start group status for the Colorado participants inherited from the HSPV study and thus could not do a follow-up comparative analyses of these two approaches in

Colorado.) Lessening of the curriculum model's impact (e.g., due to the teachers' less-extensive preservice training) may have lessened the overall impact of Head Start in Colorado.

Although the LTBHS study's findings provide encouragement that Head Start can provide long-term benefits, the study suggests that more research is needed to understand why Head Start programs in certain localities and with certain groups appear to provide greater benefits. The findings of the LTBHS study regarding mediating factors—gender, race/ethnicity, family educational and socioeconomic background, and community poverty levels and opportunities—demonstrate the influential role such factors can play in modifying Head Start's impact.

Finally, the statistical models used by the LTBHS study's analysts could be even more effective in adjusting for initial Head Start versus No–Head Start socioeconomic differences if future quasi-experimental studies would make fuller assessments of mediating factors. However, such adjustment would not be needed in those Head Start studies that employ random assignment to study groups, except in some cases (e.g., when final program and control groups differ in the rate of study participants agreeing to enroll or stay in the study).

IMPLICATIONS AND RECOMMENDATIONS

This section looks at today's Head Start programs, their expanding directions, and the related research, and in this context, considers the implications of the LTBHS study findings.

The LTBHS study's findings on Model Head Start's benefits suggest one major theme for Head Start—*exploring the potential of effectively implementing high-quality preschool programs* such as the High/Scope approach (see Hohmann & Weikart, 1995) or other child-development approaches (see Epstein, Schweinhart, & McAdoo, 1996; Stebbins et al., 1977). Here, as in the High/Scope Preschool Curriculum Comparison study, the findings of higher educational achievement and less criminal activity for the LTBHS study's Model Head Start group suggest the important contributions that high-quality preschool programs can make to participants' lives. They are contributions that last from early childhood into young adulthood.

The limitations of the LTBHS study's quasi-experimental design and methodology, along with the potential of its employed analysis methods to overcome these limitations, suggest another major theme for Head Start—*exploring the potential of high-quality research studies to demonstrate the full impact of Head Start, to investigate what constitutes quality in Head Start programs, and to promote quality in Head Start programs.* We next discuss these two basic themes with respect to directions for Head Start programs and research.

What Do the LTBHS Study's Findings Mean for Head Start Programs?

The study findings show that Head Start programs of the 1970s had some likely long-term benefits. Since that time, in numerous ways, Head Start has made strides toward promoting program quality, which should further enhance its long-term benefits. The following discussion describes the status of Head Start programs today, relating it to LTBHS study findings and other relevant research.

Today's Head Start: Serving a larger number and a greater diversity

According to the *Head Start 1998 Statistical Fact Sheet* (U.S. Department of Health and Human Services, 1998b), from 1971–83 Head Start served about 400,000 children (and families) per year. This figure had increased to about 540,000 children by 1990, to about 620,000 children by 1992, and to over 793,800 children by 1997. It has been estimated that Head Start currently enrolls 30% to 35% of 3- to 4-year-old children from families with incomes at or below the Federal poverty threshold (U.S. General Accounting Office, 1997, p. 17).

Enrollment today includes an ever-increasing number of bilingual children, some of them from families who have recently immigrated to the United States (Phillips & Cabrera, 1996). Across all Head Start programs in the nation, there are children of various ethnic/racial groups enrolled in Head Start: 36% of enrolled children are African Americans, 31% are Whites, 26% are Hispanic Americans, 4% are Native Americans, and 3% are Asian Americans; approximately 13% have disabilities (U.S. Department of Health and Human Services, 1998). The LTBHS study's Head Start participants include a somewhat comparable range of ethnic groups (although a larger proportion of African Americans), with 54% African Americans, 19% Hispanic Americans, and 27% Whites/others ("others" mainly refers to several Asian Americans and Native Americans).

Today's Head Start: Focusing on enhancing quality

A national focus on quality in Head Start programs was initiated in the late 1970s with the adoption of the national Head Start Program Performance Standards, which were revised in 1996 (U.S. Federal Register, November 1996). Based on the recommendations of the Advisory Committee on Head Start Quality and Expansion (U.S. Department of Health and Human Services, 1993), funds were provided to improve the quality of programs in various ways—by enhancing staff development and compensation, by promoting children's health and nutrition, by developing improved methods of assessment, and by enhancing relationships with schools and communities.

The LTBHS study's findings of some long-term benefits from using a model, high-quality early childhood program underscore the importance of Head Start's focus on improving program quality. These LTBHS study findings, based on the comparison of Model versus Regular Head Start, were similar to those of other studies involving the same High/Scope preschool model or other high-quality preschool programs (e.g., Campbell & Ramey, 1995; Lally et al., 1988; Schweinhart et al., 1993; Schweinhart & Weikart, 1997). Specifically, the LTBHS study's Model versus Regular Head Start findings suggest that using a model curriculum—such as the High/Scope approach or some other child development, theory-based approach—may provide long-term benefits in terms of educational achievement and positive life-adjustment (i.e., less involvement in criminal activity).

In the Head Start program in general, and in the High/Scope model program, parent-participation is an important component. Additionally, ever since its early years, the Head Start program has provided health services to children (Zigler & Muenchow, 1992) as well as other supportive services to families. Thus, other potential benefits of the overall Head Start model (e.g., child health, improvement of parental skills, family economic development), though they were not a major focus of the LTBHS study, continue to be a major feature of Head Start and its quality-enhancement efforts.

Research suggests that staff development that effectively transfers learning from model or high-quality Head Start programs to new or marginal programs should be a strong Head Start program development direction. The HSPV study documented that the Head Start teachers had been trained in 1970 to acceptable levels of High/Scope curriculum implementation (Lukas & Wohlleb, 1972; Smith, 1973), particularly at the Florida site, where the LTBHS study subsequently found some significant lasting positive effects. The High/Scope Perry Preschool study provided an *intensive* program with staff who were trained by local experts in a child development, theory-based curriculum. Like a number of other studies of model, high-quality preschool programs for low-income children (e.g., from the Consortium for Longitudinal Studies, 1983), such *intensive* program studies have typically found enduring rather than only immediate positive results.

Today's Head Start: Building communication/ partnerships with schools and communities

In addition to the benefits found for the LTBHS study's Model Head Start groups, other significant Head Start benefits were found for subgroups according to gender, race/ethnicity, or site (specifically, certain benefits found at the Florida site were not found at the Colorado site). This suggests that Head Start programs should continue to focus on improving their understanding of the characteristics and challenges of the diverse groups in their communities. This is consistent with the newest Head Start Program Performance Standards' emphasis on the importance of building communication and partnerships with the diverse communities that Head Start serves. Studies such as the National Head Start-Public School Transition Study are examining ways to extend Head Start's early effects as children proceed into their local schools. For example, based on the challenges faced by the LTBHS investigators in tracking down former Head Start participants, it appears that Head Start agencies had not typically maintained records noting which schools their children had enrolled in. Head Start agencies may find it useful to maintain such records so they can communicate and work with teachers at those schools on ways to prepare children for their school experiences.

What Do the LTBHS Study's Findings Mean for Head Start Research?

The LTBHS study, along with some previous studies, points out the methodological challenges involved in selecting measures that will provide valid and reliable estimates of Head Start's short- and long-term effects. The following recommendations for future Head Start research are based on the overall Head Start knowledge base, which now also includes the methodology and findings from the LTBHS study. (The knowledge base prior to the LTBHS study was reviewed in Chapter 2).

Continue to conduct studies of Head Start's contributions to children's development

Looking at the findings to date of studies of Head Start and of other preschool programs for economically disadvantaged children, we can find a pattern of program benefits over a range of variables—improved school-readiness, higher early-grade achievement, better school-adjustment, fewer grade retentions, and fewer referrals to special education. However, there are discrepancies—some studies found limited or no effects. It must be kept in mind that quasi-experimental studies, as opposed to random-selection studies, are likely to underestimate the true program effects, as demonstrated in the LTBHS study and discussed as early as 1975 by Donald Campbell and Robert Boruch, and more recently by others (e.g., Hebbeler, 1985; Lee & Loeb, 1995). These investigators have pointed out that discrepancies in findings across Head Start studies may reflect such specific methodological limitations as differences in the socioeconomic backgrounds of the Head Start and comparison groups (as in the LTBHS study). Underestimation of Head Start's effects is especially likely in a quasi-experimental study where Head Start programs serve the most economically disadvantaged children and families, and somewhat less disadvantaged children and families then necessarily make up the study's No–Head Start, comparison group.

However, as suggested in Chapter 2, the differences in the outcomes across Head Start studies thus far should cause us neither to deny the effectiveness of Head Start nor to give up on Head Start research. Rather, we should continue to conduct increasingly sound and appropriate studies to assess Head Start's true impact. The LTBHS study, within limits, demonstrated the potential of a quasi-experimental follow-up study to estimate long-term Head Start effects. By even more adequately assessing study groups' initial background differences and employing the analytic methods of the LTBHS study, future quasi-experimental studies should find that Head Start benefits children's development and thus their future life-outcomes.

Study Head Start programs that attempt to achieve enduring as well as immediate benefits

To effectively link children's early Head Start experiences to their achievement and competence throughout the school years and in the community, we need studies that assess the effects of using high-quality program approaches *and* studies that examine how to establish Head Start partnerships with the schools and communities. Successful programs most likely have early, or immediate, effects because they are providing certain kinds of stimulation or experiences that the family, school, or surrounding environment is not providing. According to an analysis of compensatory programs in the schools by Campbell and Frey (1970) and others (Barnett, 1992; Powell & Roane, 1996), unless comparable types and levels of experiences are provided throughout the elementary and later school years, we should not expect early effects to persist.

This is not to say that early effects do not exist or that all early effects (e.g., persistence in learning) diminish. For example, the High/Scope Perry Preschool study found that although some specific early effects (e.g., improved intelligence test scores) diminished over time, the study's preschool group, compared with its no-preschool group, went on to ultimately have higher educational attainment and better life-adjustment as young adults. Several other high-quality early childhood programs have been shown to help children to be more resilient throughout their schooling (see Campbell & Ramey, 1995; Lally et al., 1988; Schweinhart et al., 1993). Providing educational experiences of comparable quality throughout all school years might produce even greater benefits than those already found for children who had a good early childhood program.

Conduct studies that focus on components of quality-oriented program development

The next stage of Head Start research should focus on identifying the indicators of quality in Head Start programs, as recommended in the *Beyond the Blueprint* report (Phillips & Cabrera, 1996; also see Frede, 1995; Gomby et al., 1995). The knowledge base includes many examples of benefits from high-quality early childhood programs (especially demonstration programs). Now we can add to these examples the indications of long-term benefits for Model versus Regular Head Start in the LTBHS study. Research that focuses on Head Start program quality would contribute to the goals of obtaining greater consistency in short-term effects across Head Start programs and of obtaining lasting effects. How to make quality adaptations affordable is a related direction for research.

One of the challenges in conducting this program-quality research is to develop measures of quality that are consistent with the goals of Head Start and with the knowledge base on child and family development. In this endeavor, the *Beyond the Blueprint* report suggests that research measures include qualitative as well as quantitative assessment. For example, program-quality research might include observational assessments, case studies, and other types of accounts that would provide empirical analysis of how well programs meet standards of quality or how identified high-quality programs support children and families and provide for staff supervision and development. The current Head Start Quality Research Centers are identifying and developing a range of assessment measures focusing on program quality and child development outcomes.

Employ a diversity of research designs and methodological approaches

To develop a better understanding of factors that may affect Head Start's impact, we need to use a variety of research approaches. The LTBHS study findings and the Chapter 2 review of the Head Start knowledge base examined potential explanations for the discrepancies between Head Start's short- and long-term effects and between the findings across Head Start studies. As indicated throughout this book, many of the methodological problems of the LTBHS study and of other Head Start studies appear to have resulted in underestimations of Head Start's benefits. *This is reason for continuing to support Head Start and also for expanding research focused on the continuing development of Head Start.* We suggest that this expansion include the following variety of research approaches:

Studies that acknowledge the diversity in programs and communities. To expand the Head Start knowledge base, assessment approaches need to take into account the diversity across Head Start families, programs, and communities. As proposed in the *Beyond the Blueprint* report (Phillips & Cabrera, 1996), we need to develop valid quantitative and qualitative assessments of the range that exists in program quality (e.g., Bryant, Lau, Burchinal, & Sparling, 1994; Frede, 1995); in child and family development in various cultural groups (e.g., Washington & Oyemade, 1987); and in background characteristics of the children, families, and settings included in Head Start programs, neighborhoods, and communities (e.g., Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993; Rog, 1994).

Small-scale, high-quality studies. As cited in the *Beyond the Blueprint* report and supported by the two High/Scope longitudinal studies (Schweinhart et al., 1993; Schweinhart & Weikart, 1997), the Head Start knowledge base needs high-quality, small-scale studies. Such studies, especially when they are quasi-experimental, require a detailed assessment of their study groups' background (including site) characteristics. Where feasible, small-scale studies that employ random assignment to study groups (*intensive* studies) should be used to provide estimates of what young people's lives would be like without Head Start. Within small-scale studies that are limited to Head Start participants, random assignment to variations of program approaches is also a viable, informative research direction (Campbell & Boruch, 1975; Schweinhart & Weikart, 1997). Even in small-scale, intensive Head Start studies, however, there can be drawbacks: They may encounter problematic attrition of study group participants, their no–Head Start control groups may enroll in alternative early childhood programs with variations in types of child care, or their small sample-sizes may mean lack of representativeness or generalizability.

Representative, large-scale studies. Several investigators have argued for the value of large-scale, nationally representative (*extensive*) studies (Barnett, 1992; Duncan, 1991; McCall & Appelbaum, 1991). From such studies, we can initially learn more about the population served and its characteristics, and this knowledge can complement the small-scale studies. Furthermore, we can conduct secondary analyses with randomly selected populations that are either representative of the nation as a whole or representative of subgroups (e.g., on the National Longitudinal Survey of Youth, see Chase-Lansdale, Mott, Brooks-Gunn, & Phillips, 1991). Because conducting studies that initially use random assignment to program or no-program groups is not always feasible or ethical, Duncan (1991) pointed out that conducting secondary analyses on randomly chosen subsamples of large-scale studies allows a range of investigators from various fields to significantly expand the knowledge base.

Both Duncan (1991) and Barnett (1992) have pointed out that a given study will not anticipate all the lessons that can be derived from a large, representative set of data. For example, Lee and colleagues (Lee, Brooks-Gunn, & Schnurr, 1988; Lee et al., 1990) reanalyzed the data from the Educational Testing Service's Head Start Longitudinal Study, which had apparently achieved considerable representativeness of Head Start programs. They found positive Head Start effects on children's general cognitive/analytic ability as well as on other indicators of school adjustment.

Also, Lee and Loeb (1995) conducted secondary analyses on data from the National Education Longitudinal Study of 1988 (NELS:88), a representative study of the nation's eighth-graders (approximately 25,000 eighth-graders participated, 14% of whom had attended Head Start). Their analyses showed that Head Start children are subsequently educated in schools of significantly lower quality, as compared with the schools of comparison-group children. Currie and Thomas (1995), conducting secondary analyses of the National Longitudinal Survey of Youth (NLSY), compared children who had attended Head Start with their siblings who had not. Their analyses revealed that the persistence of significant gains in test scores across the elementary and middle school years is related to the children's race/ethnicity, which may be related to the schools they attend. Specifically, African American Head Start children's early test score gains, compared with those of African American no–Head Start children, were more likely to diminish as the children progressed through school.

Moving in the direction of large-scale, representative studies, the U.S. Administration for Children and Families has recently initiated a nationally representative, large-scale study of Head Start families and children. The Head Start Family and Child Experiences Survey (FACES) investigators are collaborating with the Head Start Quality Research Centers as the latter conduct small-scale intensive studies of how quality features of Head Start programs are related to program effectiveness.

Longitudinal studies of Head Start participants. Many of the small- and large-scale studies we have been describing should be longitudinal. As pointed out by Walberg (1992), a well-designed small-scale study with random assignment to treatment and control groups, if it is long-term, can examine patterns of developmental progress more extensively.

Many investigators have pointed to the advantages of following the same program participants over time and circumstances (e.g., Friedman, Haywood, & Livesey, 1994). For example, Duncan, Coe, Corcoran, Hill, and Hoffman (1984), in the Panel Study of Income Dynamics longitudinal studies, were able to discover that many families, instead of remaining in poverty, move in and out of poverty over the years. The longitudinal High/Scope Perry Preschool Study (Schweinhart et al., 1993) was able to discover long-term positive effects of preschool for low-income children. The longitudinal research by Werner and Smith (1977, 1982) revealed the resiliency of some youngsters who, inspired by nonparental role models, eventually overcame impoverished, difficult family conditions. The long-term studies by Elder (1974) and by Brooks-Gunn, Phelps, and Elder (1991) revealed the effects that major events like the Great Depression had on children and families.

Numerous investigators have specified the advantages as well as the disadvantages of longitudinal studies (e.g., Dunst & Trivette, 1994; Friedman et al., 1994; Walberg, 1992). For studying Head Start children and families, longitudinal studies offer many advantages: They are designed to track the same individuals over time and examine an implicit causal sequence; they provide estimates of effects for programs that may have a delayed impact; they allow examination of mediating effects; they allow for secondary analyses using improved statistical models; and they provide solid grounds for theory building and testing.

Still, longitudinal studies present major challenges: tracking study participants over several years, employing measures that will not become outdated over the course of the study, minimizing study participant attrition, and dealing with problems of missing data on assessments. Yet, if Head Start researchers were to employ multiple small- and large-scale longitudinal studies with commonality as well as variation in methods of assessment, this mix of studies would provide a solid basis for discerning patterns of findings across studies, like those patterns found across the various short-term studies of Head Start. In line with this idea, in 1997 the U.S. Department of Education and the National Center for Education Statistics, along with other agencies, launched the Early Childhood Longitudinal Study (ECLS) Program.

SUMMARY OF LESSONS LEARNED

The LTBHS study contended with the many challenges inherent in a quasi-experimental follow-up study: identifying, locating, and recruiting over 600 study participants from study sites in Florida and Colorado 17 years after their early childhood years; assessing their original neighborhoods' socioeconomic characteristics; and identifying appropriate statistical analyses. Also, the study was necessarily retrospective, because the data on initial family background had to be collected during the follow-up interviews, when the study participants were young adults. Earlier chapters discussed and described the study's methodological issues and decisions in considerable detail.

Many of this study's findings as well as its methodological and statistical approaches offer the following lessons for future quasi-experimental, longitudinal follow-up studies of Head Start:

- Focus on research questions that extend beyond the question of whether Head Start has positive effects; attempt to identify the specific Head Start program characteristics and experiences that provide various types of benefits. For example, determine the benefits of employing a child development, theory-based curriculum model or the benefits of using particular methods to enhance staff training and development.
- Study population characteristics of Head Start communities; use U.S. Bureau of the Census and other population data sources.
- Utilize valid, reliable instrumentation to assess family socioeconomic characteristics and to gather other kinds of background data.
- Conduct planned, longitudinal studies of Head Start graduates; work with community agencies and the schools to facilitate locating Head Start graduates at early and later stages in their post–Head Start lives.
- In analyses of data from quasi-experimental designs, look for sources of selection bias and other types of hidden bias, and employ statistical models to adjust for study groups' initial differences.

CONCLUSION

Certain features of the Long-Term Benefits of Head Start study's methodology limit our ability to reach specific conclusions about the long-term benefits of Head Start. Overall, the data analyses that adjusted for the study groups' initial socioeconomic differences produced a pattern of significant differences and nonsignificant trends similar to the patterns of findings of the High/Scope Perry Preschool study and the High/Scope Preschool Curriculum Comparison study. Several analyses revealed significant differences between study subgroups at the Florida site: The adjusted data analyses found that, compared with No–Head Start females, Head Start females had a greater rate of high school/GED completion *and a lower incidence of criminal* arrests. *Both adjusted and unadjusted data analyses revealed that, compared with the Regular Head Start group, the Model (High/Scope) Head Start group averaged a higher* GPA *in elementary, middle, and high school as well as fewer criminal* convictions.

The findings of this study—along with the findings of the High/Scope Perry Preschool study, the High/Scope Preschool Curriculum Comparison study, and others (e.g., Lally et al., 1988)—suggest that investments in high-quality early childhood education for children from economically disadvantaged backgrounds will increase their later educational attainment and lessen their negative life-outcomes (e.g., criminal activity). Despite the limitations of its quasi-experimental, retrospective design, the LTBHS study provides a pattern of findings, a system of data collection methods, and an array of statistical analysis models that provide useful guidance and direction to future Head Start research.

Appendix A

Directories and Other Resources Used in Locating Study Participants

To locate potential LTBHS study participants (referred to simply as study participants hereafter in this appendix), a number of directories, computer databases, and other resources were useful. Although today there are extensive computer databases that can readily be used to locate individuals, at the time of the study's data collection, the only existing large databases for locating individuals were ones that required investigators to know study participants' Social Security numbers. Because no Social Security numbers were available from the records used in the LTBHS study, the LTBHS staff could not use the large databases available at the time. However, it should be noted that even when investigators do have Social Security numbers at a study's outset, using some types of databases that can be accessed by Social Security number often requires that one have permission for entering a name with a Social Security number—e.g., in specific credit-check computer searches. In the LTBHS study, credit records could not be used for another reason: When study participants are from low-income backgrounds and are in the age-19 to age-22 bracket, they often use no credit cards. Thus, lacking other alternatives, we used the following resources to obtain study participants' family addresses and telephone numbers; in cases where a study participant did not have a telephone, contact by telephone with a neighbor or family member became a part of the search strategy.

DIRECTORY INFORMATION FROM SCHOOLS

In general, directory information was available from the records of the Head Start agencies and the school systems. Many schools required that their own staff be the ones to gather the information. Some school districts provided computer printouts with directory information from the time of a student's entry up to his/her last year of enrollment. The key identifying information that was collected included

- *Name of study participant*—Full, proper name of the study participant and any name changes due to adoption or marriage (the new name to be recorded in our file, along with the original, using a consistent order)

- *Other identifying information*—The study participant's gender, birth date, and ethnicity (when there was any indication)
- *School-entry data*—The actual date of initial (kindergarten or first-grade) registration
- *Most recent school attended*—The name of the school the study participant had most recently attended, graduated from, or transferred to
- *Parental addresses and telephone numbers*—The parents' location information, both at the time of school entry and at the time the school last recorded it
- *Yearbooks and class photographs*—Any yearbooks and class photographs available from the schools that could be useful for a study participant's identification

PUBLIC RECORDS

Several kinds of public records were used to find parents and study participants. We reviewed, for example, county marriage, birth, and voter registration records (the birth records were those of study participants' children). Most valuable to us, after the Head Start and school records, was each state's driver license database, which was accessible by a computer search when we paid a small fee per individual (in some cases, the fee was waived); we also needed to supply the full name and birth date of the individual. However, where feasible, names of individuals (sometimes along with birth dates) were often sufficient to yield the most recent address and often the telephone number of parents or study participants. When we knew that a family or the study participant had moved to a certain state, a computer search of the driver license database was conducted in that state.

DIRECTORIES

In addition to standard telephone directories, several other kinds of directories were useful. The most useful were the city directories used by businesses, which are obtainable by contacting the directory publishers or a city's Chamber of Commerce. Individuals can be located in these directories by street as well as by name, and some directories also provide a person's place of employment or work telephone number. Thus, we reviewed the streets in these directories for the names of study participants' family members or relatives and for their places of employment. When a telephone number was no longer in service or no longer belonged to the same party, local and regional telephone operators helped by providing a likely new telephone number. Many times the local project coordinators and locators would finally locate a study participant by calling everyone in a given area who had the same surname, thus perhaps finding a relative of the study participant.

U.S. POST OFFICE

For study participants with a known address but no telephone number, the U.S. Post Office personnel recommended that we write "Forwarding address requested" on a mailing. Thus, when there had been an address change, the new address was recorded on the mailing, and it was returned to us. The new location information was then used to pursue the study participant at that address.

INSTITUTIONAL SEARCHES

Searches were conducted in prisons, at military bases, on college campuses, at churches, and with major employers. To determine if some of our potential study participants were in prison, local study coordinators worked with state and federal prison administrative staff, who used their computer systems to search for names on our lists. The local study coordinators then arranged to contact those former students found to be in prison; thus several study interviews were conducted in prison or just after the study participant was released from prison. For a large-scale national study, major central searches in both the prison and military systems may be feasible. Military locator assistance on military bases was used in states where the study staff learned that study participants were stationed. In a few cases, the local study coordinators also contacted mental health institutions and hospitals when it was learned that a study participant might be located there. Sometimes a community's human services department was able to indicate whether a study participant was living in the area, if the person had applied for assistance in finding employment or was enrolled in a publicly funded work program.

The local study coordinators sometimes also checked with major employers—e.g., factories, businesses, and local colleges—to locate a particular study participant or parent. Churches in the major neighborhoods were sometimes able to provide a recent address of a church member. When study staff learned that a potential study participant was attending a distant college, they contacted the college to get an updated telephone number.

PUBLIC ANNOUNCEMENTS

Announcements about the LTBHS study (describing "a follow-up study of former students from the school district") with the local study coordinator's telephone number were posted at community churches, at schools, at Head Start centers, and on community bulletin boards. We also made general announcements of the study on the radio and in various local newspapers. In Colorado, the study was announced on the radio both in English and in Spanish. This publicity mainly helped to create a community awareness, to legitimize the study, and to assure acceptance of the study and the interviewers. Although the announcements helped us in locating some individuals, most people who called to inquire about the study were found not to be among those identified as eligible for the study. Whenever locators and interviewers were in neighborhoods for the first time, the local police department was notified so that the study could be verified as legitimate by anyone with a concern. Similarly, contact persons at Head Start and in the school systems were identified for interviewer-verification purposes. All interviewers carried official High/Scope Educational Research Foundation identification.

NEIGHBORHOOD CANVASSING

Local study coordinators, locators, and interviewers canvassed key neighborhoods, gave out information about the study, and showed residents the lists of former school district students not yet found. For interviews that were difficult to schedule—which was the case for approximately 5% to 10% of the study participants—a Head Start or a school employee who was well-known in the community accompanied the local study coordinator to provide introductions and assurance about the study. Also, since many low-income families do not have a telephone, some neighborhood or specific-home visits were necessary.

Appendix B

The Study's Core Variables

THE DATA SOURCES FOR THE VARIABLES

Described extensively in Chapters 3 and 4, the LTBHS study's data sources are also briefly described here as an overview of the study's database and as a preface to the presentation of the core variable specifications.

The Follow-up Interview

The study interview (at age 22) inquired primarily about the study participant's family background and present life situation, including the study participant's education and employment and whether there was any use of public assistance or any criminal activity. The interview questions and variables were designed using methodology previously developed and employed in High/Scope's longitudinal studies (e.g., Berreuta-Clement, et al.,1984; Oden et al., 1992; Schweinhart et al., 1993) and in consultation with researchers from the University of Michigan's Institute for Social Research (ISR).

The Adult APL Survey

The APL Survey instrument administered to the study participants at the time of the interview is a 40-item, multiple-choice literacy assessment covering areas pertaining to utilizing community resources, gaining employment, being a consumer, and understanding government and laws.

U.S. Bureau of the Census Data

Socioeconomic tract-level data were derived from the 1970 U.S. Bureau of the Census data, coded and entered for each study participant according to the family's original address in 1970–1971. (At one site, the 1970 population density did not require tract-level data; thus, 1980 data were substituted to provide an estimate.) Yu Xie, one of the independent data analysts, derived the tract-level data for each study site to examine the site differences—the percentage of poverty, the young-adult school dropout rate, and the percentage of female-headed households. These variables typically are considered to be major indexes of area poverty.

Educational Records Data

Study participants' records of years of schooling, course grades, special education placement, test scores, behavioral reports, suspensions, expulsions, and teacher comments from kindergarten through grade 12, as well as study participants' postsecondary courses and credits (including college and training for specialist certification), were collected at both study sites. These records covered 1971 through 1991.

Public Assistance Records Data

To determine whether each study participant had ever utilized public assistance, an initial computer search was conducted at the two sites' county human service or welfare agencies. For those study participants who had used public assistance and also had given permission for us to collect data from their records, we coded use of AFDC, food stamps, or Medicaid. Public assistance records data were not collected outside the jurisdiction area of the two study sites. However, study participants' own reports of their use of public assistance up to and at the time of the interview provided an additional data source for the public assistance core composite variable. Access to the records on study participants' Medicaid usage was very limited and thus not included in the construction of the core variables for data analysis.

Criminal Records Data

To determine whether each study participant ever had a criminal file, an initial computer search was conducted at the two sites' local and county courts. For any study participant who did have a criminal file, the data collected included the number and types of adult arrests, trials, convictions, and sentences (including time in prison and on probation). Although it was not feasible to collect out-of-state criminal data, the criminal activity core composite variables included the self-reported criminal activity from the interview, meaning arrests and convictions that study participants indicated as occurring before age 18 (so they were not in the adult criminal records) or occurring after age 18 outside the area of the study site.

DATA CODING

Standard coding and computer-entry procedures were employed using coding data manuals developed for the study. Documentation of the data-coding procedures was prepared for the data analysts. As is often the case, some variables required developing specially designed coding systems (e.g., such as creating a common K–12 grading system that was applied across multiple school districts' grading systems). Other variables required coding based on several data sources. One example is the occupations data derived from the study interview, which were supplemented with information derived from additional data sources—from the study participants' original school-entry registration forms (collected from the schools) and from the initial family-contact forms (collected by the local study coordinators, locators, or interviewers during their initial contacts with families to locate study participants). In cases of discrepancies regarding specific occupations, we selected the occupation that was identified on the school record, if available. Otherwise, parents were simply coded as having been employed at the time.

For both parents' and study participants' occupations, each job was coded using the U.S. Bureau of the Census codes in 1970 (see U.S. Bureau of the Census, 1971). Over 400 U.S. Census three-digit codes that refer to occupational titles were employed as codes for the occupational data. On the advice of data analyst Yu Xie, some occupations were moved to other classification levels to reflect typical educational training requirements or income status. Subsequently, Xie recoded the data into Duncan's Socioeconomic Index codes ranging from 1–100 (O. D. Duncan, 1961).

SELECTION AND FORMATION OF THE STUDY'S CORE COVARIATES AND OUTCOME VARIABLES

Based on specific predictions from the High/Scope Perry Preschool study findings (Schweinhart et al., 1993), the *core covariates* and *outcome variables* (as described in Chapter 4) were selected from the LTBHS study's larger database to address the study's major research questions.

The core set of variables were derived from the LTBHS study's database in consultation with the study's two data analysts. Some variables (e.g., gender) were derived from only one data source and thus were not recoded, although additional data sources were often used to confirm the status of such variables. Composite variables were either derived from several variables within one data source (e.g., from several variables that pertained to data in educational records) or from two or more data sources (e.g., from interview data and various records data). Other variables were recoded to accommodate the requirements of the various analytic approaches. Thus, a given variable might be set up to be continuous or categorical (binary or multilevel) to be employed as required for specific data analyses (see McCall & Appelbaum, 1991). The two data analysts also sometimes combined several levels of a given variable for analysis purposes.

The core covariates and outcome variables were selected and prepared for the data analyses as specified in Tables B.1 and B.2 on pp. 128–134. (Note that in these tables, the variables and their specific labels are capitalized and in bold type.)

Table B.1: Description and Specifications of Core Covariates

Variable	Sources & Notes	Variable Labels & Coding
Study Participant's Group Status:		
• For Florida and Colorado: **HEAD START** or **NO–HEAD START** • For Florida Head Start: **MODEL HEAD START** or **REGULAR HEAD START**	The assignment of study participants to groups was determined using 1970 Head Start records and fall 1971 school records. For the No–Head Start group assignment, the school records provided addresses of potential study participants who lived in neighborhoods that had been identified for Head Start recruitment or in nearby high-poverty areas. Three interviewed study participants were eliminated from the final sample because records indicated they had been out of the age-range (from 1964–67), and two others were eliminated because their Head Start or No–Head Start status could not be confirmed. In Florida, Head Start study participants were also identified as having been in either Model (High/Scope) or Regular Head Start, according to the records of the earlier, Head Start Planned Variation study (Smith, 1973).	**HS_D:** 0 = No–Head Start group 1 = Head Start group **HS_PV:** 0 = No–Head Start group 1 = Regular Head Start group 2 = Model Head Start group
Study Participant's Individual Characteristics:		
• GENDER	Originally derived from the school records and confirmed by each study participant at time of interview.	**GENDER:** 1 = male 2 = female
• RACIAL/ETHNIC MEMBERSHIP	Originally derived from the follow-up interview and recoded into two categories at each site; a few study participants indicated Asian or Native American or other group memberships and were recoded to White/other (see Table 3.2).	**RACETHN:** 0 = White/other 1 = African American (in Florida); Hispanic American (in Colorado)
• AGE AT INTERVIEW	Converted to months by computer. Final variable derived by computer from the study participant's birth date (verified in the interview) and the date of the interview.	**AGE:** total number of months at time of interview
• KINDERGARTEN PARTICIPATION	Whether or not a study participant attended kindergarten was determined from the school records and not from the interview, because there was confusion in study participants' recall as to whether they had participated in Head Start or kindergarten. (Because kindergarten had just begun to be offered in Florida, it was often referred to as preschool.) Study participants' parents confirmed that there were conflicts in recall. Thus, in cases with no records available for confirmation, kindergarten attendance was coded as 'not known.'	**FINALK:** 1 = no, never attended kindergarten 2 = yes, attended kindergarten 9 = kindergarten attendance not known.

Table B.1: Description and Specifications of Core Covariates (cont.)

Variable	Sources & Notes	Variable Labels & Coding
Study Participant's Family Background:		
• NUMBER OF SIBLINGS	Derived from the interview data by combining the reported number of older and younger siblings throughout study participants' lives up to the follow-up interview. Includes brothers and sisters, stepbrothers and stepsisters (if included by study participants as siblings).	**SIBS:** total number of siblings.
• HOUSEHOLD SIZE/STRUCTURE	Derived from the follow-up interview data. Includes total number of children (siblings, cousins, others) and adults living in the study participant's original household in 1970–71. Includes stepmother or stepfather (if study participant reported one living in the household). One-parent and two-parent households were indicated by the total number of parents in the household.	**TOTCHLDR:** total number of children in original household **TOTADULTS:** total number of adults in original household **TOTHSHLD:** total number of people in the original household **TOTPRNTS:** total number of parents in original household
• MOTHER'S EDUCATION	Derived from the follow-up interview from study participants' responses to questions asking them to recall mother's highest level of education at the time of the study participant's own school entry. (Substituted stepmother's education if the study participant indicated a stepmother and no other mother.) The highest year of education ranged from elementary school through postsecondary level. These data were used to form a categorical variable with a 4-point scale (covering from "less than high school" through "some postsecondary education"). Owing to recall problems, there was considerable missing data, so the category "unknown" was one point on this scale.	**EDULMO:** mother's highest year of education completed **EDMO_C:** 1 = unknown 2 = less than high school 3 = high school graduate/GED 4 = some postsecondary education
• FATHER'S EDUCATION	Derived from the follow-up interview from study participants' responses to questions asking them to recall father's highest level of education at the time of the study participant's own school entry. (Substituted stepfather's education if the study participant indicated stepfather and no other father.) The highest year of education ranged from elementary school through postsecondary level. These data were used to form a categorical variable with a 4-point scale covering from "less than high school" through "some postsecondary education." Owing to recall problems, there was considerable missing data, so the category "unknown" was one point on this scale.	**EDULFA:** father's highest year of education completed **EDFA_C:** 1 = unknown 2 = less than high school 3 = high school graduate/GED 4 = some postsecondary education

Table B.1: Description and Specifications of Core Covariates (cont.)

Variable	Sources & Notes	Variable Labels & Coding
Study Participant's Family Background (cont.)		
• **MOTHER'S EMPLOYMENT**	Derived from three sources: (a) interview responses to questions asking study participants to recall whether or not the mother was working and the type of job (if working) about the time of study participant's own school entry (at age 4 to 6 years), (b) registration records at study participant's school entry, (c) brief interview conducted with parents to locate study participants. In case of discrepancies, records-data indications of employment were used for the coding. The types of jobs were coded according to U.S. Bureau of the Census occupational classification codes (3-digit) and subsequently recoded from 1–100 according to Duncan's Socioeconomic Index (SEI). Whether or not the mother ever worked was also coded, even if no job type was indicated.	**MSEI:** The Duncan's SEI, continuous variables from 1 to 100. 777 = not working 989 = working, no occupation specified 999 = unknown **MAWORK2:** 1 = never worked 2 = worked at some point during study participant's preschool, elementary, or teen years 9 = unknown
• **FATHER'S EMPLOYMENT**	Derived from three sources: (a) interview responses to questions asking study participants to recall whether or not the father was working and the type of job (if working) about the time of study participant's own school entry (at age 4 to 6 years), (b) registration records at study participant's school entry, (c) brief interview conducted with parents to locate study participants. In case of discrepancies, records-data indications of employment were used for the coding. The types of jobs were coded according to U.S. Bureau of the Census occupational codes (3-digit) and subsequently recoded from 1–100 according to Duncan's Socioeconomic Index (SEI). Whether or not the father ever worked was also coded, even if no job type was indicated.	**PSEI:** The Duncan's SEI, continuous variables from 1 to 100. 777 = not working 989 = working, no occupation specified 999 = unknown **PAWORK2:** 1 = never worked 2 = worked at some point during study participant's preschool, elementary, or teen years 9 = unknown

Table B.1: Description and Specifications of Core Covariates (cont.)

Variable	Sources & Notes	Variable Labels & Coding
Study Participant's Family Background (cont.)		
• FAMILY WELFARE ASSISTANCE	Derived from the study participants' responses to follow-up interview questions about family use of welfare when the study participants were in the elementary school and the teen years. In analyses, these two variables were sometimes combined to indicate if the family ever or never had welfare assistance.	**USWLFEL:** 0 = never 1 = occasionally 2 = frequently 9 = unknown **USWLFRTN:** 0 = never 1 = occasionally 2 = frequently 9 = unknown **USEWLFREV:** 0 = never 1 = occasionally 2 = frequently 9 = unknown
• SITE TRACT-LEVEL CHARACTERISTICS	Derived from U.S. Bureau of the Census (1970 & 1980) data for the study's census tracts, which were identified from the study participants' 1970–71 addresses.	**TRACT:** tract number **POV** = % of persons living in poverty **FHH** = % of households headed by a female **HSD** = % of youths who dropped out of high school

Table B.2: Description and Specifications of Core Outcome Variables

Variable	Sources & Notes	Variable Labels & Coding
Study Participant's Education:		
• EDUCATIONAL ATTAINMENT	The highest year of education completed was derived from two data sources: (a) follow-up interview data from study participant's report on own completed education and (b) records from elementary, middle, and high school and postsecondary records, obtained subsequently as applicable. For this composite variable, the study participant's highest reported year of education was accepted. A second, categorical variable on study participant's level of education was constructed using a 4-point scale that covered "less than high school" through "some postsecondary education." The category "unknown" was one point on this scale. A postsecondary/college education variable with a 2-point scale was also created.	**EDULNEW:** highest year of education completed **ED_C:** 1 = unknown 2 = less than high school 3 = high school graduate/GED 4 = some postsecondary education **EDUPSNEW:** 0 = no college 1 = some postsecondary education
• GRADE POINT AVERAGE	Derived from the school records data for kindergarten through high school. Kindergarten courses were typically graded as satisfactory or unsatisfactory or not at all and were thus not included in the final composite variables. Grades were first coded into a common grading system, which was based on the school records documentation of the meaning of each school's grading system. The common system was: 1 = A+; 2 = A; 3 = A-; 4 = B+; 5 = B; 6 = B-; 7 = C+; 8 = C; 9 = C-; 10 = D+; 11 = D; 12 = D-; 15 = F. For analysis purposes, these data were recoded into a briefer scale: 4.3 = A+ (very excellent); A = 4; A- = 3.7; and so on, to 0 = F (failure). For each study participant, these grade values were then averaged over groups of three grades to arrive at a GPA for elementary school, a GPA for middle school, and a GPA for high school.	**GPA1** = mean grade point average in elementary school (grades 1–4) **GPA2** = mean grade point average in middle school (grades 5–8) **GPA3** = mean grade point average in high school (grades 9–12)
• YEARS IN EDUCABLE MENTAL IMPAIRMENT (EMI) CLASSES	Derived from school records indicating that the study participant had been in EMI classes (for children with lower-than-average intellectual levels, *not* emotional impairment, learning disabilities, or speech/ language impairment). The variable was derived from a count for each year enrolled in EMI classes, ranging from 0 years to potentially 12 years. For some analyses, a binary variable was formed, coded 1 if the subject was ever classified as EMI and 0 otherwise.	**EMI_YRS:** total years in special education classes for children with educable mental impairment (0 to 12 years)

BEST COPY AVAILABLE

Table B.2: Description and Specifications of Core Outcome Variables (cont.)

Variable	Sources & Notes	Variable Labels & Coding
Study Participant's Education (cont.):		
• ADULT APL SURVEY SCORE	Derived from the Adult APL Survey scores (American College Testing Program, 1976) summing over 40 functional literacy assessment items (one point for each correct answer).	**APLSCORE:** APL total score (0 to 40).
Study Participant's Life Situation/Economic Status:		
• EMPLOYMENT/ SCHOOL-ENROLLMENT STATUS	Derived from the study participants' responses to a series of follow-up interview questions on employment and postsecondary education enrollment. A composite variable was formed from two variables: *employment status* pertained to whether or not employed at some point during the interview year; *school-enrollment status* pertained to whether or not enrolled in school (includes GED courses postsecondary courses, and college) at some point during the interview year.	**EMPYSCH (nominal codes, not a scale):** 1 = employed, but not in school 2 = not employed, but enrolled in school 3 = employed *and* in school 4 = not employed *and* not in school **EMPLSCH:** 0 = not in school *and* not employed 1 = in school or employed, or both
• TEEN-PARENT STATUS	Derived from the follow-up interview data from study participants' responses to a series of questions asking for dates of their children's births (if any). From children's birth dates, a study participant's own birth date, and the date of the interview, we determined how old the study participant was at each child's birth (if any). The composite variable indicated whether or not a study participant had ever been a parent when less than age 20.	**TEENMAPA:** 1 = yes, had child born when less than age 20 2 = no, did not have child born when less than age 20 (may have had a child when age 20 or older)
• PUBLIC ASSISTANCE	Derived primarily from data obtained from a computer search of county welfare records to find any study participants who had received public assistance (AFDC) during the target year (1989-90) when records were made available at both sites. Those data were combined with data from study participants' responses to interview questions about whether they were receiving various types of public assistance. The resulting composite variable indicated whether they had ever received public assistance.	**RWELCOMP:** 0 = never used public assistance 1 = used public assistance

Table B.2: Description and Specifications of Core Outcome Variables (cont.)

Variable	Sources & Notes	Variable Labels & Coding
Study Participants' Criminal Activity:		
• **ARRESTS/ CONVICTIONS**	Derived predominantly from a computer search for each study participant (to determine subjects who had criminal records in the site's local and county courts) and from subsequent examination of records. Misdemeanor and felony arrest incidents and criminal convictions were combined for each variable. The self-report arrests and convictions data from the interview were also examined. For the composite variable, if a subject's total number of self-reported arrests and convictions was greater than the number obtained from the records, the greater total was used in place of the court records total. Only a few self-reported incidents were not found in the court records. Thus, the self-report data that was included consisted predominantly of those arrests or convictions that occurred at or before age 18 or in other locations. Overall, nearly every self-reported arrest or conviction that was added was found to have occurred *outside the county.*	**TINCDCOM:** total arrests **TCONVCOM:** total convictions

Appendix C

Original Comparability of Study Groups—Additional Tables and Map

Table C.1: Study Group Background Comparability on Household Size/Structure

	Florida				Colorado			
Variable	Head Start (*N* = 160)		No–Head Start (*N* = 264)		Head Start (*N* = 130)		No–Head Start (*N* = 68)	
Household size	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
No. of children in household	4.5**	2.5	3.9	2.5	3.8***	2.2	2.7	1.5
No. of adults in household	2.0*	0.9	2.2	1.5	2.0	1.3	2.1	0.8
Household structure	*N*	%	*N*	%	*N*	%	*N*	%
One-parent	49	30.6	70	26.5	23	17.7	12	17.6
Two-parent	102	63.8	178	67.4	103	79.2	54	79.4
Other	9	5.6	16	6.1	4	3.1	2	2.9

Note. *M* = mean. *SD* = standard deviation. "Other" refers to households with relatives other than parents or to foster-parent households. There were no significant within-site differences between study groups for the proportion of families that were two-parent.
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table C.2: Study Group Background Comparability on Mother's and Father's Education

	Florida		Colorado	
Variable	Head Start (*N* = 160)	No–Head Start (*N* = 264)	Head Start (*N* = 130)	No–Head Start (*N* = 68)
Mother's highest educational level				
Less than high school	43 (26.9%)	71 (26.9%)	73 (56.2%)	12 (17.6%)
High school graduate/GED	63 (39.4%)	107 (40.5%)	27 (20.8%)	32 (47.1%)
Some postsecondary education	20 (12.5%)	43 (16.3%)	14 (10.8%)	14 (20.6%)
Unknown	34 (21.3%)	43 (16.3%)	16 (12.3%)	10 (14.7%)
Father's highest educational level				
Less than high school	44 (27.5%)	49 (18.6%)	60 (46.2%)	23 (33.8%)
High school graduate/GED	43 (26.9%)	87 (33.0%)	26 (20.0%)	24 (35.3%)
Some postsecondary education	12 (7.5%)	31 (11.7%)	7 (5.4%)	11 (16.2%)
Unknown	61 (38.1%)	97 (36.7%)	37 (28.5%)	10 (14.7%)

Table C.3: Study Group Background Comparability on Mother's and Father's Employment

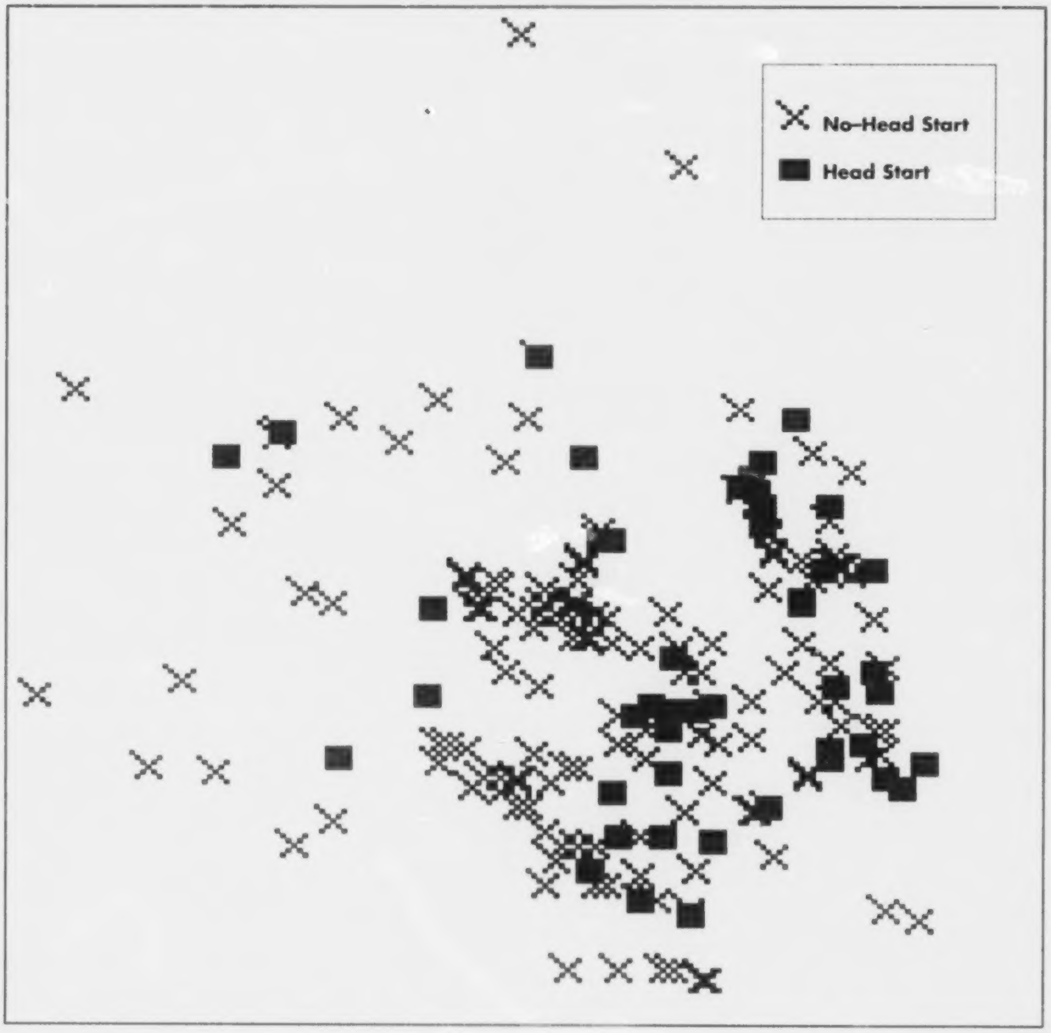
	Florida		Colorado	
Variable	**Head Start (*N* = 160)**	**No-Head Start (*N* = 264)**	**Head Start (*N* = 130)**	**No-Head Start (*N* = 68)**
Mother's employment				
Number (%) of mothers with job type identified	85 (53.1%)	124 (47.0%)	56 (43.1%)	26 (38.2%)
SEI range of job types	7.0–64.0	7.0–71.0	6.3–65.0	6.1–63.2
Median SEI rating	13.7	17.0	16.0	17.1
Mean SEI rating (*SD*)	21.2 (17.3)	24.5 (17.0)	23.4 (18.7)	31.4 (22.8)
Father's employment				
Number (%) of fathers with job type identified	99 (61.9%)	175 (66.3%)	104 (80.0%)	58 (85.3%)
SEI range of job types	4.1–77.0	4.1–92.3	6.3–62.3	6.1–84.0
Median SEI rating	15.1	17.0	13.9	16.4
Mean SEI rating (*SD*)	19.2 (16.0)	21.7 (16.1)	15.2 (10.9)	22.2 (18.9)

Note. *SD* = standard deviation; SEI = Duncan's Socioeconomic Index ratings of job types.

Table C.4: Study Group Background Comparability on Family Welfare Assistance

	Florida		Colorado	
Variable	**Head Start (*N* = 160)**	**No-Head Start (*N* = 264)**	**Head Start (*N* = 130)**	**No-Head Start (*N* = 68)**
Frequency of family welfare assistance (elementary school years)				
Never	73 (45.6%)	155 (58.7%)	62 (47.7%)	45 (66.2%)
Occasionally	32 (20.0%)	38 (14.4%)	35 (26.9%)	10 (14.7%)
Frequently	52 (32.5%)	64 (24.2%)	31 (23.8%)	10 (14.7%)
Unknown	3 (1.9%)	7 (2.7%)	2 (1.5%)	3 (4.4%)

Map of Subjects' Addresses by Head Start Status, Florida Site, Area C, Urban Regular Head Start



Appendix D

Assessment of the Long-Term Benefits of Head Start

Yu Xie
Professor of Sociology
University of Michigan

In this chapter, I present results from statistical analyses of a data set collected by the High/Scope Foundation for the Long-Term Benefits of Head Start study (hereafter the LTBHS study).

The data set consists of two subsamples, one from a site consisting of two urban and two rural communities in Florida and another from a site that is an urban community of Colorado. Although the two subsamples can be pooled to gain statistical precision, careful inspection of the data reveals that the two subsamples are different in some fundamental ways. For example, the predominant non–White minority in the Colorado subsample is Hispanic American, whereas the predominant non–White minority in the Florida subsample is African American. Furthermore, Head Start appears to have recruited subjects from more severely disadvantaged families in Colorado than in Florida (Table D.1, p. 140). As a conservative strategy, I treat the two subsamples separately and do not pool information across the two sources. That is, for the statistical results reported here, analysis was conducted separately for the Florida and the Colorado subsamples.

THE DATA

The data collection procedure is described elsewhere and will not be repeated here. Overall, the information collected from different sources appears to be of high quality and to be carefully coded. Before proceeding to data analysis, it is imperative to highlight two features of the LTBHS study design, because these features constrain data analysis and ultimately render causal inferences problematic.

The first important feature is that the assignment of subjects to treatment (i.e., the Head Start group) and control (i.e., the No–Head Start group) was not random. From teachers at local Head Start centers, we learned that conscious but unsystematic efforts were made in these locales to recruit children from poor families. For example, literature was distributed at shopping centers close to poor neighborhoods. In areas with few poor preschool-aged children, virtually all of them were recruited, leaving no comparable subjects for comparison. In areas that had numerous poor preschool-aged children (such as Florida's area C), recruitment was stopped

Table D.1: Descriptive Statistics of Outcome Variables and Covariates

	Florida Site				Colorado Site			
	No-Head Start		Head Start		No-Head Start		Head Start	
Variable	Mean/%	SD	Mean/%	SD	Mean/%	SD	Mean/%	SD
Outcome variable								
High school completion (%)	86.0		81.9		86.8		73.1	
Some postsecondary education (%)	49.6		40.0		54.4		26.9	
Incidence of EMI (%)	5.3		6.3		4.4		4.6	
GPA in grades 1–4	2.64	0.56	2.74	0.62	2.83	0.39	2.64	0.38
GPA in grades 5–8	2.26	0.63	2.25	0.74	2.69	0.73	2.24	0.68
GPA in grades 9–12	1.98	0.76	2.01	0.84	2.29	0.98	1.71	0.94
Adult APL Survey score	28.97	6.45	26.51	7.63	29.90	7.46	24.72	8.51
Arrest incidence (%)	28.0		28.1		52.9		55.4	
Conviction incidence (%)	19.7		18.1		36.8		46.2	
Receipt of public assistance (%)	32.6		43.1		32.4		46.9	
Covariate								
Race (% non-White)	80.3		78.1		32.4		73.1	
Sex (% female)	50.4		49.4		48.5		44.6	
Mother's education (%)								
Less than high school	26.9		26.9		17.6		56.2	
High school	40.5		39.4		47.1		20.8	
Some postsecondary	16.3		12.5		20.6		10.8	
Unknown	16.3		21.3		14.7		12.3	
Father's education (%)								
Less than high school	18.6		27.5		33.8		46.2	
High school	33.0		26.9		35.3		20.0	
Some postsecondary	11.7		7.5		16.2		5.4	
Unknown	36.7		38.1		14.7		28.5	
Family's welfare use (%)								
Never	58.7		45.6		66.2		47.7	
Occasional	14.4		20.0		14.7		26.9	
Frequent	24.2		32.5		14.7		23.8	
Unknown	2.7		1.9		4.4		1.5	

BEST COPY AVAILABLE

after target numbers were reached. Given the Head Start recruitment practices, construction of a comparable control group was a very difficult, if not impossible, task.

The LTBHS research staff approached this task by assembling the control group from the same communities (defined as the same school districts or neighboring school districts when the number of eligible students to be included in the No–Head Start group was insufficient). When potential candidates for the control group were numerous at a site area (such as Florida's area C), random selection was used to recruit subjects for the control group. Given that Head Start is targeted at children from poor families, it would seem probable that the control group came from families of better socioeconomic status. Thus, parental socioeconomic status (SES) is likely to be negatively related to participation in a Head Start program. Indeed, this pattern is apparent in the descriptive statistics reported in Table D.1 and is more pronounced for the Colorado subsample than for the Florida subsample.

The second important feature is that early childhood information was collected retrospectively rather than prospectively. Crucial information about family parental SES during the subject's childhood (at ages 4 to 11) was obtained from a face-to-face interview during the subject's early adulthood (between ages 19 and 22). It is possible that a subject's knowledge about his or her parents during the subject's childhood was factually inaccurate. Also, either lacking recollection or desiring to appear "normal," a subject could have distorted information during the interview.

METHODOLOGICAL ISSUES

One major methodological difficulty is the lack of assurance that the No–Head Start (hereafter abbreviated as No–HS) subjects are comparable to the Head Start (hereafter abbreviated as HS) subjects in such background characteristics as family income, parental education, and other family resources. This problem is further compounded by the lack of rich and reliable variables measuring family SES and other background information in the data set. While parental education was measured retrospectively, family income was not measured. Income can only be crudely approximated from a response to a question about the frequency with which the respondent's family received welfare assistance during the respondent's elementary school years. However, such retrospective measures are prone to recall errors. Variables measuring other forms of family resources, such as possessions, parenting practices, and parents' mental health conditions, are unavailable.

The potential noncomparability between the HS and No–HS groups poses a serious methodological threat to my analysis of the data. Surely the problem is rooted in the first feature of the study design, the fact that assignment to treatment and control was not random and thus the data were observational rather than experimental. If the subjects had been randomly assigned to the treatment and control groups as they would be in an experiment, systematic differences between the treatment and control groups would have been minimized to chance alone. A successful example of an experimental design is the well-known High/Scope Perry Preschool study (Schweinhart, Barnes, & Weikart, 1993).

As pointed out by Manski and Garfinkel (1992), experimental designs suffer from shortcomings that are often overlooked. For example, researchers and policymakers are rarely concerned with the impact of feedback and changing environments when results from an experimental study are extended to the whole population. As a complement, Manski and Garfinkel recommended the continued use of the "structural" approach, i.e., modeling causal processes statistically based on observational data. The LTBHS researchers were in part motivated by the objective to draw causal inferences through use of observational data. Unfortunately, the LTBHS study's data set does not contain rich enough background information for a true "structural" analysis. Some naive and unrealistic assumptions will have to be invoked. Thus, results reported here should be interpreted very cautiously.

I will illustrate the problem more formally, borrowing notations from Manski (1995). This approach is well known in the statistics literature commonly associated with Rubin (1974). Note that we observe the outcomes of treatment only for those subjects who participated in Head Start programs; we observe the outcomes of control for those subjects who did not participate in Head Start programs. Let z be a binary variable denoting the status of participation in Head Start ($z = 0$ for no participation, and $z = 1$ for participation). If a child participated, we observe the outcome of participation in a later time and denote it by y_1. However, we have no way of observing the outcome of nonparticipation, denoted by y_0, which exists conceptually nonetheless. Likewise, for a child who did not participate in Head Start, we only observe y_0 but not y_1. Say that we are only concerned with the expectations of y_0 and y_1, denoted by $E(y_0)$ and $E(y_1)$.[1] Unobservability means that we only observe $E(y_1|z = 1)$ but not $E(y_1|z = 0)$, and only observe $E(y_0|z = 0)$ but not $E(y_0|z = 1)$. Conceptually at least, when referring to program effects of Head Start, we mean the comparison in outcome variables between the unconditional expectations $E(y_1)$ and $E(y_0)$.

Note the relationships expressed in equations 1A and 1B:

$$E(y_1) = E(y_1|z = 0)\ P(z=0) + E(y_1|z = 1)\ P(z=1) \tag{1A}$$

$$E(y_0) = E(y_0|z = 0)\ P(z=0) + E(y_0|z = 1)\ P(z=1). \tag{1B}$$

In equations 1A and 1B, $P(z = 0)$, $P(z = 1)$, $E(y_1|z = 1)$, and $E(y_0|z = 0)$ are observable. $E(y_1|z = 0)$ and $E(y_0|z = 1)$ are unobservable. Without further strong assumptions, equations 1A and 1B are identifiable only in terms of rather broad bounds for binary dependent variables (Manski, 1995, p. 41).

---

[1] We can be interested in any characteristic of functions $P(y_0)$ and $P(y_1)$. For a more general treatment, see Manski (1995).

One convenient but overly unrealistic assumption is that selection into program participation (z) is exogenous (or "strongly ignorable") conditional on measured covariates. We denote observed covariates by x. This assumption is tantamount to equations 2A and 2B:

$$E(y_1|x) = E(y_1|x, z = 0) = E(y_1|x, z = 1) \tag{2A}$$

$$E(y_0|x) = E(y_0|x, z = 0) = E(y_0|x, z = 1). \tag{2B}$$

In essence, equations 2A and 2B assume that, conditional on observed covariates, the underlying expectations governing the unobservables [$E(y_1|x, z = 0)$ and $E(y_0|x, z = 1)$] are the same as those for their observable counterparts [respectively, ($E(y_1|x, z = 1)$ and $E(y_0|x, z = 0)$]. This assumption undergirds all the regression results presented in this report. While the assumption is justifiable in a true experimental design, which randomly assigns subjects into the two groups ($z = 0$, $z = 1$), it is unlikely to hold true in such observational data as in the LTBHS study. Thus, the effects of Head Start estimated under the assumption are likely to be biased.

It is important to emphasize that an inadequate control for SES characteristics tends to cause downward (i.e., negative) biases on the estimated effects of Head Start, because SES is likely to be correlated negatively with Head Start participation but positively with such desirable outcomes as educational achievement, low likelihood of committing crime, and economic independence. In other words, without adequate control for differences in SES between the HS group and the No–HS group, statistical analyses based on the experimental model (i.e., equations 2A and 2B) are likely to significantly underestimate the program effects of Head Start. Thus, it is plausible that the lack of true comparability between the treatment group and control group in the LTBHS study may contribute to the seemingly anomalous finding that the estimated effects of Head Start appear to be harmful.

The possibility that the HS and No–HS groups may be incomparable is further demonstrated in geographic patterns of the original households of the subjects' families. My use of the geographic information involved two steps. First, I geo-coded the subjects' addresses and plotted them onto digital maps through a Geographic Information System (GIS). Then, I linked to the sampled data the relevant contextual variables computed from the U.S. Bureau of the Census data at the tract level. Through examining the maps, it became clear that the families of the HS subjects were geographically concentrated in districts different from those where families of the No–HS subjects resided. To protect the identity of the sites, I suppressed the original maps created by the GIS in this report. The results for the contextual variables are presented in Table D.2 on p. 144. In general, compared with the No–HS subjects, the HS subjects lived in districts marked by higher incidence of poverty, school dropout, and female-headed households (as measured by aggregate statistics from the 1970 or 1980 U.S. Census).[2]

---

[2] Information from the 1980 U.S. Census was used only for one county, which was untracted in 1970

Table D.2: U.S. Bureau of the Census Tract-Level Characteristics

Site Location/HS Status	Poverty Rate		Young-Adult Dropout Rate		Female-Headed Household Rate		
	Mean %	SD	Mean %	SD	Mean %	SD	N
Colorado site							
No-Head Start	18.21	6.94	14.72	8.15	14.39	10.88	72
Head Start	21.23	7.37	17.07	9.68	18.12	12.62	127
Combined	20.14	7.34	16.22	9.20	16.77	12.12	199
Florida site, area A (urban)							
No-Head Start	7.44	2.65	9.22	3.70	16.22	5.54	9
Model Head Start	15.67	10.64	11.18	6.76	19.62	12.47	39
Combined	14.13	10.16	10.81	6.31	18.98	11.52	48
Florida site, area B (rural)							
No-Head Start	23.92	12.32	16.42	9.38	22.11	14.15	36
Model Head Start	30.97	10.82	15.46	7.46	30.06	12.13	35
Combined	27.39	12.06	15.94	8.44	26.03	13.70	71
Florida site, area C (urban)							
No-Head Start	34.81	12.53	23.21	7.56	26.16	9.67	195
Regular Head Start	32.14	14.10	20.20	8.51	24.55	10.07	71
Combined	34.09	12.99	22.40	7.92	25.73	9.78	266
Florida site, area D (rural)							
No-Head Start	19.29	8.51	17.75	8.04	9.29	4.24	24
Regular Head Start	20.12	8.62	20.62	8.08	10.41	3.37	17
Combined	19.63	8.46	18.94	8.08	9.76	3.90	41
Total sample							
No-Head Start	28.24	13.72	19.90	8.83	21.73	11.70	336
Head Start	24.27	11.87	17.06	9.11	20.89	12.51	289
Combined	26.41	13.04	18.58	9.06	21.34	12.08	625

Note: Site tract-level characteristics were computed from the 1970 U.S. Census except for the Florida site's areas A and B, for which data from the 1980 U.S. Census were used.

We approach the noncomparability problem between the HS group and the No–HS group through measured covariates (i.e., x). The better our measured covariates capture all relevant factors that determine later outcomes *and* vary systematically between the HS group and the No–HS group, the less risk there is for a systematic bias. That is, the better our x measures, the more plausible the assumption implied by equations 2A and 2B becomes. For this purpose, we include *race* (African American in Florida, or Hispanic American in Colorado, versus White/other in both states), *sex* (male versus female), *mother's education* (less than high school, high school, some postsecondary, and unknown), *father's education* (less than high school, high school, some postsecondary, and unknown), and *family's welfare use* (never, occasional, frequent, unknown) during the subject's elementary school years. These explanatory variables form the basis for my baseline model. Descriptive statistics for these covariates are reported in Table D.1 on p. 140 by site (by area within site for Florida) and by Head Start participation status.

In earlier analyses, I also included a few other variables that are available in the data set: *number of siblings, number of parents, mother's* and *father's employment status, mother's employment status when the subject was 4–6 years old,* and *mother's* and *father's occupation when the subject was 4–6 years old* (measured by Duncan's Socioeconomic Index). After experimenting with them, I decided against the use of these variables. My decision was based on three reasons: (1) The variables do not contribute any additional explanatory power beyond the base line model; the estimated coefficients in the multivariate regressions are unstable and uninterpretable in almost all situations. (2) Whether or not I include the variables, it does not change the substantive results. I suspect that these variables perform poorly either because they were seriously contaminated by measurement error or because there is not enough variation in them for our sampled subjects. (3) Some of the variables are highly correlated with other variables already included in the baseline model. For example, the number of parents is correlated with the "unknown" categories for *mother's education* and *father's education*. That is, the subjects were more likely to have nonresponse for a parent's education when the parent was absent from home. It would be a mistake to assume that these missing values (the "unknowns") were missing at random and to use the listwise deletion method to treat them. Therefore, I take a conservative approach by creating a unique category for them.

I control for systematic differences between the HS group and the No–HS group through the addition of the covariates in multivariate regressions. This solution is rather standard and straightforward. The straightforwardness of our approach in large part derives from the observability of *some* (but not all) key characteristics. If we think that the subjects in the HS group come from lower SES families than do the subjects in the No–HS group after this control, our statistical model *reduces* (but does not eliminate) biases from our estimation of program effects. Therefore, the statistical results that follow need to be carefully interpreted.

The origin of the LTBHS study was the Head Start Planned Variation study, which was reported to reveal differences in outcomes between children who attended Regular Head Start programs and children who attended Model Head Start programs using the High/Scope curriculum model. The research design for this purpose relied on regional variations in program implementation at the Florida site: Model Head Start programs were implemented in site areas A and B, whereas Regular Head Start programs were implemented in site areas C and D. Had there been no systematic regional variation net of rural/urban (which, in this study, is controlled), this design would have allowed a direct comparison of the long-term benefits of participation in Model Head Start with those of participation in Regular Head Start.

Careful examination of the data, however, convinced me that such a comparison should best be avoided. Regional variation in measurements for outcome variables is too large to be ignored, and this variation is seriously confounded with the true effects of program variation. For example, the Florida site's areas A and C were selected as urban areas for the comparison of Model Head Start versus Regular Head Start; statistics calculated from decennial U.S. censuses, reported in Table D.2, suggest that the subjects in area C lived in poorer neighborhoods than did the subjects in area A. If the subjects who attended Model Head Start in area A outperformed the subjects who attended Regular Head Start in area C, we are not in a position to uniquely attribute the difference to either program variation or regional variation. Let me further illustrate the issue of confounding with a more concrete example of the *grade point average (GPA)* variable, on which Model Head Start subjects outperformed Regular Head Start subjects in the LTBHS data. Table D.3 presents the data on this variable.

The top panel (designated as "Aggregated") in Table D.3 shows that the Model (High/Scope) Head Start subjects have noticeably higher *GPAs* than do either the Regular Head Start subjects or even the No–Head Start subjects, for all three measures of the variable (grades 1–4, 5–8, and 9–12). For example, in grades 1–4, *GPA* for the Model Head Start subjects averages 3.15, compared with 2.38 for the Regular Head Start subjects and with 2.64 for the No–Head Start subjects. Although one may be tempted to conclude from the comparison that Model Head Start was the underlying cause for the differences, further analyses reported in subsequent panels of Table D.3 suggest that regional variation may account for the observed differences. When statistics for the areas are reported in separate panels in Table D.3, for each area the within-area average difference between Head Start (whether Regular or Model) subjects and No–Head Start subjects is essentially nil. The pattern exhibited in the top ("Aggregated") panel can be explained as a result of the following two facts: (1) The subjects in areas A and B *all* received better *GPAs* (3.07 and 3.25, on average, respectively) than the subjects in areas C and D did (2.47 and 2.46, on average, respectively). (2) The Model Head Start subjects were recruited from areas A and B, and the Regular Head Start subjects from areas C and D.

Table D.3: Confounding of Regional Variation With Program Variation—Grade Point Average (GPA) and Crime Outcomes in Florida

	GPA Grades 1–4		GPA Grades 5–8		GPA Grades 9–12				
Site Location/HS Status	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**% Ever Arrested**	**% Ever Convicted**	**N**
Aggregated									
No-Head Start	2.64	0.56	2.26	0.63	1.98	0.76	28.0%	19.7%	264
Regular Head Start	2.38	0.59	2.00	0.72	1.75	0.77	30.2%	20.9%	86
Model Head Start	3.15	0.32	2.53	0.66	2.31	0.82	25.7%	14.9%	74
Area A (urban)									
No-Head Start	2.98	0.15	2.59	0.39	2.22	0.48	37.5%	12.5%	8
Model Head Start	3.08	0.23	2.25	0.70	1.99	0.85	30.8%	15.4%	39
Combined	3.07	0.22	2.31	0.67	2.03	0.79	31.9%	14.9%	47
Area B (rural)									
No-Head Start	3.28	0.39	2.85	0.53	2.62	0.75	30.6%	25.0%	36
Model Head Start	3.21	0.38	2.82	0.47	2.59	0.69	20.0%	14.3%	35
Combined	3.25	0.39	2.84	0.50	2.61	0.72	25.4%	19.7%	71
Area C (urban)									
No-Head Start	2.51	0.49	2.12	0.58	1.83	0.69	27.0%	19.4%	196
Regular Head Start	2.35	0.58	1.99	0.67	1.75	0.74	33.3%	24.6%	69
Combined	2.47	0.52	2.08	0.61	1.81	0.70	28.7%	20.8%	265
Area D (rural)									
No-Head Start	2.41	0.57	2.23	0.68	1.99	0.79	29.2%	16.7%	24
Regular Head Start	2.54	0.64	2.06	1.02	1.73	0.91	17.7%	5.9%	17
Combined	2.46	0.59	2.18	0.79	1.90	0.83	24.4%	12.2%	41

Model Head Start seems to have a beneficial effect in reducing crime. In the seventh and eighth columns of Table D.3, the percentages of the subjects *ever arrested* and *ever convicted* are presented by site area and program status. In area B, the subjects who participated in Model Head Start have lower arrest rates and conviction rates than do the subjects who did not participate in Head Start. However, the sample size is too small to render the results statistically significant. In sum, I cannot conclude with confidence that Model Head Start had any long-term effects that are distinctive from Regular Head Start and unconfounded by site characteristics.

From analyses such as these reported in Table D.3, I became convinced of the importance of controlling for overall community differences. For the regression analyses explained next, I achieved the control through the use of a set of dummy variables. Although I also experimented with the tract-level characteristics as reported in Table D.2, the estimated coefficients of such aggregate variables are mostly insignificant and uninterpretable. I suspect that these variables are not truly comparable across locales. This leaves me with the conservative approach of saturating between-tract differences for our final models through the use of tract-specific dummy variables. In essence, the coefficients of the dummies reflect net average differences across the U.S. Bureau of the Census tracts in measured outcome variables.

With regional variation thus controlled, I conveniently pooled into a single category the subjects who had attended both Model Head Start and Regular Head Start; I then contrasted them with subjects who had not attended Head Start. In this framework, I do not ask how outcome variables vary across site areas. Rather, I ask whether the effect of Head Start is the same across site areas. That is, the differential effects of Model Head Start and Regular Head Start are expressed as the interaction effect between *area* and *Head Start status*. For all regressions reported in the following discussion, no such interaction is found to be significant. In other words, there is no evidence in the LTBHS study data to suggest that the subjects who attended Model Head Start did significantly better than the subjects who attended Regular Head Start.

STATISTICAL RESULTS

Educational Attainment

In this section, I examine the differences between the HS subjects and the No–HS subjects first regarding the likelihood of *high school completion* and then regarding the likelihood of *attainment of some postsecondary education*.

The bivariate relationship between *Head Start participation* and *high school completion* is negative. As shown in Table D.1 on p. 140, at the Florida site 86.0% of the No–HS subjects, compared with 81.9% of the HS subjects, completed high school education. At the Colorado site the contrast is 86.8% versus 73.1% in favor of the No–HS subjects. To make these bivariate results comparable with coefficients from logit regressions, we can convert the rawpercentage differences into log-odds ratios (*LOR*). Denoting y as a binary dependent variable ($y = 1$ if *high school completion*, and $y = 0$ otherwise) and inserting the observed proportions for probabilities, we obtain

$$\begin{aligned}&\log \{[P(y = 1|\text{HS})/P(y = 0|\text{HS})]/[P(y = 1|\text{No–HS})/P(y = 0|\text{No–HS})]\}\\&= \log\{(0.819/0.181)/(0.860/0.1400)\}\\&= -0.306.\end{aligned}$$

As it should, the number obtained here corresponds exactly with the coefficient of the Head Start status for the bivariate model, shown as Model 1 in Table D.4, p. 150. The comparable *LOR* for the Colorado subsample is −0.882, indicating a greater difference between the No–HS and the HS subjects (86.8% vs. 73.1%) for the Colorado subsample than for the Florida subsample.

Besides providing a base line for later models, the bivariate Model 1 has another advantage of yielding a standard test for the statistical significance of the observed bivariate difference. Constructing a z-test by taking the ratio between the estimated coefficient and the estimated standard error (*SE*), we note that the bivariate difference between HS subjects and No–HS subjects is statistically significant ($p < .05$) for the Colorado subsample ($z = -2.172$) but not for the Florida subsample ($z = -1.133$).

Containing only program status and not controlling for other factors, Model 1 suggests that the odds of *high school completion* are lower among the HS subjects than among the No–HS subjects. As discussed earlier, this bivariate relationship can be spurious, because the HS subjects came from more disadvantaged backgrounds than did the No–HS subjects. To control for such differences between the HS and the No–HS groups in background characteristics, I further include a set of variables in Model 2: *race, sex, mother's education* and *father's education* (in four categories: less than high school, high school, some postsecondary, and unknown), and *family's welfare use* (in four categories: never, occasional, frequent, and unknown). In addition, I include a set of dummy variables for unique community characteristics at the U.S. Bureau of the Census tract level in Model 2.[3] Model 2 can be written more formally as follows, letting y_i^* denote the latent dependent variable, $\boldsymbol{x}_i$ a vector for controlled covariates, z_i the binary indicator for Head Start status, for the ith subject:

$$y_i^* = \boldsymbol{x}'_i\beta + \eta z_i + \varepsilon_i \tag{3}$$

In equation 3, β is the parameter vector for controlled covariates, η the parameter for Head Start status, and ε_i the residual. If the dependent variable is continuous, y_i^* is the same as observed y_i. For continuous dependent variables, the ordinary least squares (OLS) method is used for estimation. If the dependent variable is dichotomous, as in the present case, the following

---

[3] I use the dummy variables to "saturate" differences across the communities. Thus, estimated coefficients for these dummy variables do not have substantive interpretation but simply index intertract mean differences.

Table D.4: Logit Regression Models Predicting High School Completion

Site/Variable	Model 1 Parameter	Model 1 (*SE*)	Model 2 Parameter	Model 2 (*SE*)	Model 3 Parameter	Model 3 (*SE*)
Florida (n = 424)						
Intercept	1.814	(0.176)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.306**	**(0.270)**	**0.297**	**(0.376)**	**-0.840**	**(0.518)**
Race (1 = non–White)			0.465	(0.480)	0.384	(0.493)
Sex (1 = female)			0.218	(0.340)	-0.663	(0.450)
Mother's education (excluded = less than high school)						
High school			0.565	(0.410)	0.642	(0.424)
Some postsecondary			1.360	(0.908)	1.526	(0.969)
Unknown			0.440	(0.495)	0.520	(0.505)
Father's education (excluded = less than high school)						
High school			0.623	(0.498)	0.624	(0.515)
Some postsecondary			1.880	(1.162)	1.743	(1.179)
Unknown			-0.137	(0.456)	-0.036	(0.471)
Family's welfare use (excluded = never)						
Occasional			0.425	(0.586)	0.568	(0.609)
Frequent			-1.533	(0.380)	-1.587	(0.389)
Unknown			10.450	(22.400)	11.120	(22.190)
Head Start*sex					**2.345**	**(0.743)**
L^2/DF	365.42	422	265.17	373	254.35	372
Colorado (n = 198)						
Intercept	1.880	(0.355)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.882**	**(0.406)**	**-0.049**	**(0.532)**	**-0.379**	**(0.743)**
Race (1 = non–White)			-1.679	(0.571)	-1.672	(0.572)
Sex (1 = female)			-0.161	(0.417)	-0.690	(0.902)
Mother's education (excluded = less than high school)						
High school			0.820	(0.600)	0.837	(0.605)
Some postsecondary			0.274	(0.782)	0.272	(0.785)
Unknown			-0.429	(0.616)	-0.434	(0.621)
Father's education (excluded = less than high school)						
High school			-0.125	(0.619)	-0.118	(0.621)
Some postsecondary			0.987	(1.149)	1.016	(1.148)
Unknown			-0.696	(0.540)	-0.670	(0.544)
Family's welfare use (excluded = never)						
Occasional			-0.219	(0.547)	-0.228	(0.549)
Frequent			-0.708	(0.552)	-0.709	(0.556)
Unknown			-2.602	(1.154)	-2.623	(1.146)
Head Start*sex					**0.670**	**(1.009)**
L^2/DF	204.60	196	160.22	174	159.77	173

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. $L^2 = -2$ log-likelihood; *DF* = degrees of freedom.

threshold measurement is called for:

$$y_i = 1 \text{ if } y_i^* > 0 \qquad (4)$$
$$y_i = 0 \text{ if otherwise.}$$

Assuming an independent standard logistic distribution for ε_i, parameters in equation 3 can be estimated via maximum likelihood (ML).

The ML estimates for Model 2 are reported by the middle pair of data columns in Table D.4. We observe that after controlling for the battery of covariates, for the Florida subsample, the effect of Head Start on *high school completion* is turned positive, albeit statistically insignificant at a conventional significance level. For the Colorado subsample, the coefficient of Head Start is close to zero, indicating that Head Start participation has no net effect on *high school completion.*

Most of the covariates affect *high school completion* in predictable ways. Note that we treat the problem of missing data for *mother's education, father's education,* and *family's welfare use* by creating a unique category and thus allowing respondents with missing values (a response of "unknown") on these variables to differ from other respondents on the dependent variable. Excluding them from analysis would reduce efficiency as well as introduce a bias.[5] At the Florida site, the coefficient for *race* is positive but not statistically different from zero, suggesting no net difference between African Americans and Whites/others on the independent variable. At the Colorado site, the estimated coefficient of *race* indicates that the odds of *high school completion* for Hispanic Americans are significantly lower than those for Whites/others ($LOR = -1.679$). The net sex difference on *high school completion* is nil for both subsamples. Both *mother's education* and *father's education* greatly increase the subject's odds of *high school completion* for the Florida subsample, but the effects of these covariates are statistically insignificant for the Colorado subsample. Likewise, frequent *family welfare use* significantly decreases the odds of *high school completion* for the Florida subsample but does not have the same statistical power for the Colorado subsample.

Model 2 assumes an additive effect of Head Start. From earlier studies, we know that the effect of Head Start may differ by sex. For example, the High/Scope Perry Preschool study (Schweinhart et al., 1993, pp. 54–80) found that the educational benefits of High/Scope programs are much more significant for females than for males. This interaction effect, if present, can help identify the true benefits of Head Start. Denote the dummy variable *sex* by s ($s = 1$ for female), and consider the following model:

$$y_i^* = \boldsymbol{x}'_i\beta + \alpha z_i + \eta z_i s_i + \varepsilon_i \qquad (5)$$

From our earlier discussion, we may expect a situation where $\alpha = 0$, but

---

[5] As commented earlier, it is highly plausible that missing values (the "unknowns") on these variables were not missing at random. For example, some respondents who did not report father's education probably did not know their biological fathers.

$\eta \neq 0$, which corresponds to a significant effect of Head Start for females but nil effect for males.[5]

With this reasoning, we further include the interaction between *sex* and *Head Start status* in Model 3 (see the last pair of data columns in Table D.4, p. 150), and this interaction effect is tantamount to allowing the effect of Head Start to vary between males and females. The estimated coefficient of this interaction is positive (*LOR* = 2.345) and statistically significant for the Florida subsample ($z = 3.156$, $p < 0.05$), indicating that the Head Start program has a far greater return for females than for males. The estimated interaction effect is in the same direction but does not attain statistical significance ($z = 0.664$, $p > 0.10$) for the Colorado subsample. For the Florida subsample, it appears that the effect of Head Start may be negative for males (*LOR* = −0.663) but positive for females (*LOR* = 1.682). If we invoke another strong assumption, that the treatment of Head Start cannot hurt children (Manski, 1995, pp. 44–50), the negative effect observed among males (i.e., estimated α) is likely due to residual SES differences between the HS and No–HS subjects that are omitted in our model. According to this interpretation, the effect of Head Start on *high school completion* is nil for males but highly significant for females (that is, estimated η).

I now turn to the second dependent variable for this section: the likelihood of subjects' *attainment of some postsecondary education*. Given the generally poor background of the subjects in the two subsamples, it is perhaps surprising to observe that a large fraction of them have attained some postsecondary education. As shown in Table D.1 on p. 140, 49.6% of the No–HS subjects and 40% of the HS subjects in Florida attained some postsecondary education. In Colorado the percentages are 54.4% among the No–HS subjects and 26.9% among the HS subjects. Two details about the dependent measurement are worth noting here. First, the measurement is essentially attendance at, not completion of, any postsecondary education. Second, the measurement is a composite, based on both the subject's self report during the interview and administrative records from postsecondary schools in which the subject was likely to enroll.

Similar to Table D.4, Table D.5 presents three logit models predicting the likelihood of *attainment of some postsecondary education*. Model 1 shows that the bivariate effect of Head Start participation is negative and that the magnitude of the effect is more pronounced and more statistically significant for the Colorado subsample than for the Florida subsample. This result is consistent with the descriptive statistics reported in Table D.1 on p. 140.

In Model 2, we add our standard battery of controls and reverse the sign of the coefficient of Head Start from negative to positive for the Florida subsample and change the negative coefficient of Head Start from −1.175 to −0.780 for the Colorado subsample. In general, *mother's education* and

---

[5] If the reverse situation (significant effect for males but not for females) is true, $\alpha \neq 0$, but $\eta = -\alpha$. This situation is also empirically testable.

Table D.5: Logit Regression Models Predicting Attainment of Some Postsecondary Education

Site/Variable	Model 1 Parameter	Model 1 (*SE*)	Model 2 Parameter	Model 2 (*SE*)	Model 3 Parameter	Model 3 (*SE*)
Florida (n = 424)						
Intercept	-0.015	(0.123)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.390**	**(0.203)**	**0.046**	**(0.272)**	**-0.436**	**(0.375)**
Race (1 = non-White)			0.734	(0.345)	0.724	(0.347)
Sex (1 = female)			0.518	(0.245)	0.175	(0.304)
Mother's education (excluded = less than high school)						
High school			0.601	(0.309)	0.625	(0.311)
Some postsecondary			1.253	(0.449)	1.260	(0.454)
Unknown			-1.003	(0.434)	-0.992	(0.437)
Father's education (excluded = less than high school)						
High school			-0.133	(0.349)	-0.138	(0.352)
Some postsecondary			0.855	(0.527)	0.808	(0.528)
Unknown			-0.259	(0.359)	-0.258	(0.361)
Family's welfare use (excluded = never)						
Occasional			0.023	(0.336)	0.061	(0.338)
Frequent			-0.645	(0.305)	-0.670	(0.308)
Unknown			-0.159	(0.848)	-0.155	(0.846)
Head Start*sex					**0.957**	**(0.506)**
L^2/DF	581.33	422	449.08	373	445.44	372
Colorado (n = 198)						
Intercept	0.177	(0.243)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**-1.175**	**(0.313)**	**-0.780**	**(0.417)**	**-0.715**	**(0.570)**
Race (1 = non-White)			-0.539	(0.390)	-0.542	(0.390)
Sex (1 = female)			-0.496	(0.360)	0.577	(0.609)
Mother's education (excluded = less than high school)						
High school			0.608	(0.449)	0.609	(0.449)
Some postsecondary			1.185	(0.570)	1.193	(0.573)
Unknown			-0.751	(0.767)	-0.743	(0.768)
Father's education (excluded = less than high school)						
High school			-0.054	(0.458)	0.057	(0.459)
Some postsecondary			1.153	(0.654)	1.155	(0.655)
Unknown			0.216	(0.545)	0.213	(0.545)
Family's welfare use (excluded = never)						
Occasional			0.060	(0.470)	0.063	(0.470)
Frequent			-0.482	(0.563)	-0.481	(0.564)
Unknown			-8.478	(16.980)	-8.517	(17.050)
Head Start*sex					**-0.125**	**(0.756)**
L^2/DF	245.19	196	199.64	174	199.61	173

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. L^2 = -2 log-likelihood; *DF* = degrees of freedom.

father's education have positive effects and *family's welfare use* has negative effects on *attainment of some postsecondary education.* Only when estimates are statistically insignificant does this pattern not hold true. For example, the coefficient of *father's education* being "high school" is negative with a large standard error, meaning that there is no net difference between the subjects who had fathers with "less than high school" and the subjects who had fathers with "high school" education.

The results of Model 3 reveal that for Florida the interaction effect between *Head Start status* and *sex* is significantly positive ($LOR = 0.957$, with $z = 1.891$, $p < 0.10$). This pattern is the same as the earlier result for *high school completion.* That is, the results indicate that compared with females who did not participate in Head Start, females who participated in Head Start are more likely to go on to attain postsecondary education. This interpretation is made stronger if we accept the proposition that the Head Start program does not decrease males' likelihood of *attainment of some postsecondary education.*

Intellectual Capacity and Performance

Five variables from the LTBHS data set are selected as measures of intellectual capacity and performance: *years in educable mental impairment (EMI) classes, grade point average (GPA) in grades 1–4, GPA in grades 5–8, GPA in grades 9–12,* and *total score on the Adult APL Survey.* Because whether a subject was in EMI class varied from year to year of his/her schooling, *incidence of EMI* was coded as a binary dependent variable: 1 if a subject was ever classified as having an educable mental impairment in any school year, and 0 otherwise. As with the earlier analyses for *high school completion* and *attainment of some postsecondary education,* logit models will be used for *incidence of EMI.* The other four dependent variables are continuous and will be modeled using OLS regressions.

As shown in Table D.1 on p. 140, at the bivariate level, the HS subjects had higher *incidence of EMI* than did the No–HS subjects. The contrast is 6.3% versus 5.3% in Florida, and 4.6% versus 4.4% in Colorado. Translated into *LOR,* the bivariate effect of Head Start is 0.174 for the Florida subsample and 0.047 for the Colorado subsample, and neither effect is statistically significant. (See Model 1, Table D.6.)

The introduction of control variables in Model 2 reverses the sign of the coefficient of Head Start from positive to negative, although it is not statistically different from zero for either subsample. In Model 3, I further add the interaction effect between *Head Start status* and *sex* and find it statistically insignificant ($p > 0.10$) at the conventional significance level. In sum, I am led to conclude that the evidence in the LTBHS data set does not suggest that Head Start programs have much beneficial influence on mental ability as measured by *incidence of EMI.*

The *GPA* measures were obtained from school records by averaging grades over all courses taken by a subject. To improve reliability while allowing

Table D.6: Logit Regression Models Predicting Incidence of Educable Mental Impairment (EMI)

Site/Variable	Model 1 Parameter	Model 1 (SE)	Model 2 Parameter	Model 2 (SE)	Model 3 Parameter	Model 3 (SE)
Florida (n = 424)						
Intercept	-2.882	(0.274)	(4 dummy variables for sites)			
Head Start (1 = yes)	**0.174**	**(0.426)**	**-0.265**	**(0.525)**	**-0.828**	**(0.645)**
Race (1 = non-White)			0.316	(0.624)	0.333	(0.620)
Sex (1 = female)			-1.038	(0.487)	-1.746	(0.699)
Mother's education (excluded = less than high school)						
High school			-0.993	(0.629)	-0.952	(0.636)
Some postsecondary			-7.681	(14.490)	-7.619	(14.280)
Unknown			0.491	(0.593)	0.599	(0.603)
Father's education (excluded = less than high school)						
High school			-0.248	(0.739)	-0.321	(0.744)
Some postsecondary			-0.083	(1.199)	-0.121	(1.208)
Unknown			0.324	(0.648)	0.278	(0.648)
Family's welfare use (excluded = never)						
Occasional			-0.340	(0.720)	-0.279	(0.722)
Frequent			0.604	(0.495)	0.638	(0.497)
Unknown			-8.286	(33.510)	-8.322	(33.020)
Head Start*sex					**1.610**	**(0.982)**
L^2/DF	184.29	422	154.20	408	151.40	407
Colorado (n = 198)						
Intercept	-3.076	(0.586)	-6.010	(1.823)	-5.648	(1.829)
Head Start (1 = yes)	**0.047**	**(0.719)**	**-0.023**	**(1.021)**	**-0.971**	**(1.391)**
Race (1 = non-White)			-2.444	(1.059)	-2.287	(1.08
Sex (1 = female)			0.524	(0.827)	-0.715	(1.526)
Mother's education (excluded = less than high school)						
High school			-0.767	(1.583)	-0.872	(1.653)
Some postsecondary			1.081	(1.704)	1.147	(1.713)
Unknown			-0.295	(1.149)	-0.324	(1.158)
Father's education (excluded = less than high school)						
High school			1.916	(1.366)	2.040	(1.401)
Some postsecondary			-6.711	(22.850)	-6.236	(23.660)
Unknown			2.740	(1.372)	2.820	(1.365)
Family's welfare use (excluded = never)						
Occasional			3.001	(1.430)	3.096	(1.489)
Frequent			3.688	(1.466)	3.808	(1.515)
Unknown			-5.965	(45.720)	-5.872	(44.890)
Head Start*sex					**1.813**	**(1.828)**
L^2/DF	73.22	196	46.03	185	44.99	184

Note. Census tracts are not included because of marginal zeros (i.e., no incidence of EMI within each tract). In Models 2 and 3 for Florida, sample sites are controlled. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. L^2 = -2 log-likelihood; *DF* = degrees of freedom.

for variation over time, I aggregated *GPA* over 4-year intervals: *grades 1–4*, *grades 5–8*, and *grades 9–12*. The descriptive statistics in Table D.1 (p. 140) show that the No–HS subjects generally had a higher *GPA* than did the HS subjects. The exceptions are *grades 1–4* in Florida, for which the HS subjects had a slightly higher *GPA*, and *grades 5–8 and 9–12* in Florida, for which there is virtually no difference between the HS subjects and the No–HS subjects.

Linear regressions were estimated via ordinary least squares (OLS) to predict all three measures of academic performance: *GPA in grades 1–4*, *GPA in grades 5–8*, and *GPA in grades 9–12*. The results are reported in Tables D.7–D.9. Although the tables show the estimated coefficient of Head Start to be significantly negative for the Colorado subsample, it may be an artifact attributable to the fact that the HS subjects in the Colorado subsample are severely selective with unfavorable background characteristics as compared with the No–HS subjects. For the Florida subsample, the effect of Head Start on *GPA* is mostly nil, with the exception that there appears to be an interaction effect between *Head Start status* and *sex* for *GPA in grades 9–12* (the interaction effect is 0.309, with $z = 1.919$, $p < 0.10$). Assuming no effect instead of a negative effect of Head Start for males due to omitted variables, this interaction effect suggests that Head Start's academic return is significantly positive for females.

The Adult APL survey has a minimum total score of 0 and a maximum total score of 40. On average, compared with No–HS counterparts, the HS subjects performed poorly on the test. In Florida the HS subjects' scores averaged 26.51, whereas the No–HS subjects' scores averaged 28.97. In Colorado the disparity is even greater: an average of 24.72 for the HS subjects compared with an average of 29.90 for the No–HS subjects. Regression analysis (Model 1 of Table D.10, p. 160) shows that these bivariate differences are statistically significant ($z = -3.422$ for Florida and $z = -4.035$ for Colorado, $p < 0.05$).

The inclusion of control variables in Model 2 greatly *reduces* the negative coefficient of Head Start on *Adult APL Survey scores*. For the Florida subsample, this results in the statistical insignificance of the HS and No–HS difference. For the Colorado subsample, the magnitude of the net effect is smaller (from −5.185 to −3.412) but remains statistically different from zero ($z = -2.413$, $p < 0.05$). Again, the latter result may be due to severe negative selectivity in the Colorado subsample. I also tested but found to be insignificant the interaction effect between *Head Start status* and *sex*. In testing for interactions, I found a sizable and statistically significant ($p < 0.10$) interaction effect between *Head Start status* and *race* for the Florida subsample. The results are presented as Model 3. The coefficient of Head Start is significantly negative for White/other subjects (−3.474, $p < 0.05$), but the gap between HS and No–HS narrows to nil for African American subjects ($-3.474 + 3.375 = 0.099$). If a case can be made that the negative coefficient for Whites/others is due to selectivity, it seems that participation in Head Start improves *Adult APL Survey scores* for African American subjects (by more than 3 points).

Table D.7: OLS Regression Models Predicting GPA in Grades 1–4

	Model 1		Model 2		Model 3	
Site/Variable	**Parameter**	***(SE)***	**Parameter**	***(SE)***	**Parameter**	***(SE)***
Florida (n = 350)						
Intercept	2.639	(0.040)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**0.103**	**(0.064)**	**-0.033**	**(0.056)**	**-0.018**	**(0.074)**
Race (1 = non–White)			-0.100	(0.072)	-0.099	(0.072)
Sex (1 = female)			0.274	(0.051)	0.286	(0.064)
Mother's education (excluded = less than high school)						
High school			0.134	(0.065)	0.134	(0.065)
Some postsecondary			0.346	(0.090)	0.348	(0.090)
Unknown			0.050	(0.087)	0.049	(0.088)
Father's education (excluded = less than high school)						
High school			-0.038	(0.072)	-0.038	(0.072)
Some postsecondary			0.089	(0.105)	0.089	(0.105)
Unknown			-0.091	(0.076)	-0.092	(0.076)
Family's welfare use (excluded = never)						
Occasional			0.018	(0.072)	0.016	(0.073)
Frequent			-0.026	(0.061)	-0.026	(0.062)
Unknown			-0.180	(0.173)	-0.179	(0.173)
Head Start*sex					**-0.031**	**(0.104)**
SSE/DF	117.91	348	56.65	299	56.63	298
Colorado (n = 161)						
Intercept	2.827	(0.052)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.183**	**(0.064)**	**-0.131**	**(0.079)**	**-0.156**	**(0.107)**
Race (1 = non–White)			-0.030	(0.075)	-0.027	(0.076)
Sex (1 = female)			-0.041	(0.064)	0.074	(0.112)
Mother's education (excluded = less than high school)						
High school			0.207	(0.085)	0.209	(0.085)
Some postsecondary			0.196	(0.114)	0.194	(0.114)
Unknown			-0.060	(0.104)	-0.060	(0.104)
Father's education (excluded = less than high school)						
High school			-0.026	(0.085)	0.025	(0.086)
Some postsecondary			0.198	(0.143)	0.203	(0.144)
Unknown			0.065	(0.089)	0.066	(0.089)
Family's welfare use (excluded = never)						
Occasional			0.016	(0.082)	0.014	(0.083)
Frequent			0.037	(0.092)	0.037	(0.093)
Unknown			-0.273	(0.236)	-0.270	(0.237)
Head Start*sex					**0.048**	**(0.135)**
SSE/DF	23.01	159	20.09	137	20.07	136

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. *SSE* = sum of squares error; *DF* = degrees of freedom.

Table D.8: OLS Regression Models Predicting GPA in Grades 5–8

	Model 1		Model 2		Model 3	
Site/Variable	**Parameter**	**(*SE*)**	**Parameter**	**(*SE*)**	**Parameter**	**(*SE*)**
Florida (n = 355)						
Intercept	2.261	(0.046)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.006**	**(0.074)**	**-0.018**	**(0.072)**	**-0.096**	**(0.097)**
Race (1 = non–White)			-0.167	(0.093)	-0.174	(0.093)
Sex (1 = female)			0.278	(0.065)	-0.217	(0.082)
Mother's education (excluded = less than high school)						
High school			0.091	(0.083)	0.090	(0.083)
Some postsecondary			0.168	(0.114)	0.162	(0.114)
Unknown			0.050	(0.111)	0.057	(0.111)
Father's education (excluded = less than high school)						
High school			0.054	(0.092)	0.053	(0.092)
Some postsecondary			0.438	(1.133)	0.441	(0.132)
Unknown			0.039	(0.096)	0.044	(0.096)
Family's welfare use (excluded = never)						
Occasional			-0.020	(0.092)	-0.010	(0.092)
Frequent			-0.227	(0.080)	-0.227	(0.080)
Unknown			-0.344	(0.240)	-0.344	(0.239)
Head Start*sex					**0.161**	**(0.133)**
SSE/DF	161.08	353	95.62	304	95.16	303
Colorado (n = 146)						
Intercept	2.687	(0.100)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.451**	**(0.122)**	**-0.361**	**(0.141)**	**-0.351**	**(0.182)**
Race (1 = non–White)			0.080	(0.136)	0.078	(0.138)
Sex (1 = female)			-0.011	(0.113)	0.003	(0.200)
Mother's education (excluded = less than high school)						
High school			0.188	(0.150)	0.187	(0.151)
Some postsecondary			0.158	(0.200)	0.158	(0.201)
Unknown			0.135	(0.179)	0.135	(0.180)
Father's education (excluded = less than high school)						
High school			0.201	(0.154)	0.201	(0.154)
Some postsecondary			0.330	(0.265)	0.326	(0.269)
Unknown			0.137	(0.156)	-0.137	(0.157)
Family's welfare use (excluded = never)						
Occasional			-0.006	(0.147)	-0.004	(0.148)
Frequent			-0.413	(0.158)	-0.412	(0.158)
Unknown			-1.092	(0.402)	-1.094	(0.404)
Head Start*sex					**-0.021**	**(0.245)**
SSE/DF	69.30	144	50.65	122	50.65	121

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. *SSE* = sum of squares error; *DF* = degrees of freedom.

Table D.9: OLS Regression Models Predicting GPA in Grades 9–12

	Model 1		Model 2		Model 3	
Site/Variable	**Parameter**	**(SE)**	**Parameter**	**(SE)**	**Parameter**	**(SE)**
Florida (n = 352)						
Intercept	1.980	(0.053)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**0.030**	**(0.087)**	**-0.028**	**(0.088)**	**-0.177**	**(0.117)**
Race (1 = non-White)			-0.003	(0.114)	-0.011	(0.113)
Sex (1 = female)			0.182	(0.079)	0.069	(0.098)
Mother's education (excluded = less than high school)						
High school			0.080	(0.102)	0.080	(0.102)
Some postsecondary			0.050	(0.139)	0.041	(0.139)
Unknown			-0.008	(0.136)	0.002	(0.136)
Father's education (excluded = less than high school)						
High school			0.105	(0.111)	0.106	(0.111)
Some postsecondary			0.423	(0.158)	0.421	(0.158)
Unknown			0.016	(0.117)	0.027	(0.116)
Family's welfare use (excluded = never)						
Occasional			-0.117	(0.111)	-0.097	(0.111)
Frequent			-0.285	(0.098)	-0.286	(0.098)
Unknown			-0.392	(0.271)	-0.404	(0.270)
Head Start*sex					**0.309**	**(0.161)**
SSE/DF	217.10	350	141.81	303	140.09	302
Colorado (n = 128)						
Intercept	2.295	(0.151)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.580**	**(0.182)**	**-0.482**	**(0.209)**	**-0.535**	**(0.258)**
Race (1 = non-White)			-0.272	(0.193)	-0.263	(0.195)
Sex (1 = female)			0.137	(0.163)	0.047	(0.301)
Mother's education (excluded = less than high school)						
High school			0.334	(0.218)	0.346	(0.221)
Some postsecondary			0.624	(0.272)	0.616	(0.274)
Unknown			-0.137	(0.259)	-0.132	(0.260)
Father's education (excluded = less than high school)						
High school			0.181	(0.219)	0.178	(0.220)
Some postsecondary			-0.150	(0.387)	-0.137	(0.390)
Unknown			-0.254	(0.230)	-0.253	(0.231)
Family's welfare use (excluded = never)						
Occasional			0.228	(0.205)	0.219	(0.207)
Frequent			-0.020	(0.268)	-0.027	(0.270)
Unknown			-1.577	(0.538)	-1.559	(0.542)
Head Start*sex					**0.128**	**(0.362)**
SSE/DF	114.72	126	76.65	104	76.55	103

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. *SSE* = sum of squares error; *DF* = degrees of freedom.

Table D.10: OLS Regression Models Predicting Adult APL Survey Scores

	Model 1		Model 2		Model 3	
Site/Variable	Parameter	(SE)	Parameter	(SE)	Parameter	(SE)
Florida (n = 395)						
Intercept	28.970	(0.441)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**-2.457**	**(0.718)**	**-0.801**	**(0.776)**	**-3.474**	**(1.633)**
Race (1 = non-White)			-0.297	(0.959)	-1.895	(1.286)
Sex (1 = female)			2.972	(0.685)	3.017	(0.683)
Mother's education (excluded = less than high school)						
High school			1.696	(0.905)	1.777	(0.903)
Some postsecondary			1.674	(1.227)	1.723	(1.223)
Unknown			-2.291	(1.154)	-2.172	(1.152)
Father's education (excluded = less than high school)						
High school			0.513	(1.005)	0.3562	(1.005)
Some postsecondary			1.499	(1.375)	1.224	(1.378)
Unknown			-2.044	(1.033)	-2.177	(1.031)
Family's welfare use (excluded = never)						
Occasional			-0.031	(0.980)	0.1995	(0.984)
Frequent			-0.896	(0.851)	-0.7906	(0.849)
Unknown			2.721	(2.344)	2.833	(2.337)
Head Start*race					**3.375**	**(1.817)**
SSE/DF	18793	393	13850	344	13712	343
Colorado (n = 181)						
Intercept	29.900	(1.047)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**-5.185**	**(1.285)**	**-3.412**	**(1.414)**	**-3.73**	**(1.977)**
Race (1 = non-White)			-2.279	(1.354)	-2.657	(2.130)
Sex (1 = female)			4.004	(1.165)	4.016	(1.170)
Mother's education (excluded = less than high school)						
High school			0.569	(1.511)	0.536	(1.522)
Some postsecondary			5.477	(1.932)	5.488	(1.939)
Unknown			-3.903	(2.038)	-3.905	(2.044)
Father's education (excluded = less than high school)						
High school			-0.389	(1.539)	0.358	(1.550)
Some postsecondary			0.642	(2.299)	0.5926	(2.316)
Unknown			1.412	(1.698)	1.406	(1.704)
Family's welfare use (excluded = never)						
Occasional			-1.630	(1.542)	-1.653	(1.550)
Frequent			-3.419	(1.692)	-3.410	(1.698)
Unknown			-3.429	(3.625)	-3.426	(3.636)
Head Start*race					**0.623**	**(2.702)**
SSE/DF	11960	179	8802.1	158	8799.1	157

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. *SSE* = sum of squares error; *DF* = degrees of freedom.

Crime

We use two variables to measure crime incidence, and both are composites constructed from self-reported criminal histories and from court records in the states of Florida and Colorado. The first is a binary variable, *arrest incidence,* indicating whether a subject had been arrested for any reason by 1991 (1 = yes, and 0 = otherwise). The second is another binary variable, *conviction incidence,* indicating whether a subject had been convicted of crime (1 = yes, and 0 = otherwise). Table D.1 on p. 140 gives the mean percentages of the two dependent measures by location and by Head Start status. In Florida, the crude difference in *arrest incidence* between the No–HS subjects and the HS subjects is nil (28.0% vs. 28.1%), and *conviction incidence* is slightly less for the HS subjects than for the No–HS subjects (18.1% vs. 19.7%). However, in Colorado, it appears that rates of *arrest incidence* and *conviction incidence* are higher for the HS subjects than for the No–HS subjects (55.4% vs. 52.9%, and 46.2% vs. 36.8%, respectively).

Model 1 in Tables D.11 and D.12, p. 162–163, is a bivariate logit model showing that the above observation about the bivariate relationship between *HS status* and crime—*arrest* or *conviction incidence*—is not statistically significant. In Model 2, which controls for covariates, the coefficient of Head Start is estimated to be negative, suggesting that Head Start may reduce *arrest incidence* and *conviction incidence.* This is true also for *conviction incidence* in Colorado, where the bivariate effect of Head Start of 0.388 is changed to −0.103 in the multivariate model. However, none of the negative coefficients of Head Start are statistically significant in Model 2.

Model 3 allows interaction effects between *Head Start status* and *sex.* The interaction effect is estimated to be negative in all four cases, although it is statistically significant from zero only for the arrest equation in Florida (interaction effect of −1.472, with $z = -2.265$, $p < 0.05$). The coefficient of Head Start, which refers to males only, given the interaction between *Head Start status* and *sex,* is not significantly different from zero. Thus, it seems that Head Start reduces crime only for females, not for males.

Public Assistance

From the interview and administrative records on public assistance, a binary variable was constructed to measure whether or not a subject had ever received welfare assistance (1 = yes, and 0 = otherwise). As shown in Table D.1, the rates of *receipt of public assistance* appear higher for HS subjects in both the Florida and the Colorado subsamples (43.1% vs. 32.6% in Florida, and 46.9% vs. 32.4% in Colorado). In the unit of *LOR,* the bivariate effect of Head Start is 0.451 for the Florida subsample and 0.614 for the Colorado subsample, and both effects are statistically significant ($p < 0.05$). (See Table D.13, p. 164.) This positive relationship between *Head Start status* and *receipt of public assistance* is reduced to insignificance with the control of the covariates in Model 2 for the Colorado subsample. For the Florida subsample,

Table D.11: Logit Regression Models Predicting Arrest Incidence

	Model 1		Model 2		Model 3	
Site/Variable	Parameter	(*SE*)	Parameter	(*SE*)	Parameter	(*SE*)
Florida (n = 424)						
Intercept	-0.943	(0.137)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**0.005**	**(0.222)**	**-0.207**	**(0.297)**	**0.273**	**(0.365)**
Race (1 = non–White)			0.183	(0.388)	0.199	(0.395)
Sex (1 = female)			-1.713	(0.285)	-1.262	(0.336)
Mother's education (excluded = less than high school)						
High school			-0.417	(0.349)	-0.443	(0.351)
Some postsecondary			-0.344	(0.495)	-0.289	(0.499)
Unknown			-0.380	(0.434)	-0.430	(0.438)
Father's education (excluded = less than high school)						
High school			-0.293	(0.396)	-0.304	(0.399)
Some postsecondary			-1.095	(0.634)	-1.098	(0.636)
Unknown			0.017	(0.390)	-0.029	(0.395)
Family's welfare use (excluded = never)						
Occasional			0.271	(0.371)	0.200	(0.375)
Frequent			0.564	(0.323)	0.613	(0.326)
Unknown			-2.048	(1.461)	-2.290	(1.518)
Head Start*sex					**-1.472**	**(0.650)**
L^2/*DF*	503.35	422	383.46	373	377.82	372
Colorado (n = 198)						
Intercept	0.118	(0.243)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**0.098**	**(0.300)**	**-0.449**	**(0.412)**	**-0.229**	**(0.532)**
Race (1 = non–White)			-0.027	(0.382)	-0.034	(0.382)
Sex (1 = female)			-1.543	(0.346)	-1.240	(0.578)
Mother's education (excluded = less than high school)						
High school			-0.487	(0.447)	-0.497	(0.449)
Some postsecondary			-0.877	(0.584)	-0.861	(0.584)
Unknown			-0.519	(0.626)	-0.523	(0.627)
Father's education (excluded = less than high school)						
High school			-0.261	(0.459)	-0.264	(0.459)
Some postsecondary			-0.733	(0.631)	-0.727	(0.628)
Unknown			0.763	(0.517)	0.765	(0.520)
Family's welfare use (excluded = never)						
Occasional			-0.041	(0.460)	-0.038	(0.461)
Frequent			0.102	(0.504)	0.105	(0.506)
Unknown			-1.219	(1.114)	-1.208	(1.104)
Head Start*sex					**-0.459**	**(0.710)**
L^2/*DF*	272.74	196	216.90	174	216.49	173

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. L^2 = -2 log-likelihood; *DF* = degrees of freedom.

Table D.12: Logit Regression Models Predicting Conviction Incidence

	Model 1		Model 2		Model 3	
Site/Variable	**Parameter**	**(*SE*)**	**Parameter**	**(*SE*)**	**Parameter**	**(*SE*)**
Florida (n = 424)						
Intercept	-1.405	(0.155)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**-0.103**	**(0.257)**	**-0.046**	**(0.323)**	**0.158**	**(0.395)**
Race (1 = non-White)			0.766	(0.461)	0.785	(0.465)
Sex (1 = female)			-1.469	(0.314)	-1.269	(0.380)
Mother's education (excluded = less than high school)						
High school			-0.481	(0.382)	-0.487	(0.382)
Some postsecondary			-0.594	(0.569)	-0.561	(0.572)
Unknown			-0.388	(0.460)	-0.408	(0.461)
Father's education (excluded = less than high school)						
High school			0.126	(0.449)	0.124	(0.449)
Some postsecondary			-0.283	(0.709)	-0.289	(0.710)
Unknown			0.580	(0.444)	0.561	(0.445)
Family's welfare use (excluded = never)						
Occasional			0.164	(0.414)	0.129	(0.417)
Frequent			0.425	(0.354)	0.440	(0.355)
Unknown			2.008	(1.517)	-2.120	(1.538)
Head Start*sex					**-0.599**	**(0.670)**
L^2/*DF*	413.43	422	322.86	373	322.04	372
Colorado (n = 198)						
Intercept	-0.542	(0.251)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**0.388**	**(0.307)**	**-0.103**	**(0.416)**	**0.052**	**(0.512)**
Race (1 = non-White)			0.452	(0.385)	0.444	(0.385)
Sex (1 = female)			-1.414	(0.352)	-1.145	(0.621)
Mother's education (excluded = less than high school)						
High school			-0.162	(0.448)	-0.162	(0.449)
Some postsecondary			-0.448	(0.610)	-0.434	(0.609)
Unknown			0.257	(0.588)	0.264	(0.589)
Father's education (excluded = less than high school)						
High school			-0.516	(0.473)	-0.523	(0.473)
Some postsecondary			-0.763	(0.683)	-0.764	(0.683)
Unknown			0.206	(0.484)	0.196	(0.487)
Family's welfare use (excluded = never)						
Occasional			0.034	(0.451)	0.038	(0.451)
Frequent			0.605	(0.488)	0.610	(0.489)
Unknown			0.357	(1.097)	0.349	(1.090)
Head Start*sex					**-0.388**	**(0.747)**
L^2/*DF*	268.89	196	216.82	174	216.55	173

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. L^2 = -2 log-likelihood; *DF* = degrees of freedom.

BEST COPY AVAILABLE

Table D.13: Logit Regression Models Predicting Receipt of Public Assistance

	Model 1		Model 2		Model 3	
Site/Variable	Parameter	(*SE*)	Parameter	(*SE*)	Parameter	(*SE*)
Florida (n = 424)						
Intercept	-0.727	(0.131)	(39 dummy variables for census tracts)			
Head Start (1 = yes)	**0.451**	**(0.207)**	**0.544**	**(0.293)**	**0.703**	**(0.412)**
Race (1 = non-White)			-0.274	(0.370)	-0.260	(0.371)
Sex (1 = female)			1.790	(0.272)	1.913	(0.355)
Mother's education (excluded = less than high school)						
High school			-0.657	(0.333)	-0.665	(0.334)
Some postsecondary			-0.267	(0.472)	-0.269	(0.471)
Unknown			-0.375	(0.428)	-0.387	(0.430)
Father's education (excluded = less than high school)						
High school			-0.249	(0.376)	-0.239	(0.377)
Some postsecondary			-1.904	(0.633)	-1.889	(0.632)
Unknown			0.081	(0.386)	0.082	(0.386)
Family's welfare use (excluded = never)						
Occasional			0.449	(0.373)	0.433	(0.374)
Frequent			0.993	(0.317)	0.992	(0.317)
Unknown			1.512	(0.845)	1.514	(0.849)
Head Start*sex					**-0.294**	**(0.533)**
L^2/*DF*	552.01	422	406.15	373	405.85	372
Colorado (n = 198)						
Intercept	0.738	(0.259)	(12 dummy variables for census tracts)			
Head Start (1 = yes)	**0.614**	**(0.313)**	**0.377**	**(0.429)**	**0.453**	**(0.603)**
Race (1 = non-White)			0.501	(0.394)	0.500	(0.394)
Sex (1 = female)			1.591	(0.366)	1.691	(0.670)
Mother's education (excluded = less than high school)						
High school			-0.590	(0.462)	-0.595	(0.463)
Some postsecondary			-0.079	(0.600)	-0.077	(0.601)
Unknown			1.123	(0.613)	1.126	(0.614)
Father's education (excluded = less than high school)						
High school			-0.533	(0.464)	-0.532	(0.464)
Some postsecondary			-1.225	(0.727)	-1.229	(0.728)
Unknown			-0.733	(0.518)	-0.734	(0.517)
Family's welfare use (excluded = never)						
Occasional			0.017	(0.479)	0.018	(0.478)
Frequent			0.976	(0.498)	0.976	(0.498)
Unknown			0.367	(1.035)	0.365	(1.039)
Head Start*sex					**-0.139**	**(0.771)**
L^2/*DF*	265.34	196	211.00	174	210.97	173

Note. Census tract of original residence is controlled in Models 2 and 3. The regression results were estimated under the assumption that selection into Head Start programs was exogenous conditional on the included variables. L^2 = -2 log-likelihood; *DF* = degrees of freedom.

the control variables have little impact on the coefficient of Head Start but increase its standard error to make it marginally significant according to a z-test ($z = 1.857$, $p < 0.10$). There is no evidence that the Head Start effect on females differs from that on males, because the interaction effect between *Head Start status* and *sex* is insignificant for both subsamples. In other words, Model 3 does not improve upon Model 2. These results do not lend support to the hypothesis that participation in Head Start reduces dependency on welfare.

CONCLUSION

It is a statistical truism that empirical evidence is used only to *reject*, not to confirm, hypotheses. Although my preceding analyses were designed to test the validity of many interesting hypotheses derived from earlier studies on the long-term benefits of Head Start, the exercise was focused around an implicit null hypothesis: that participation in Head Start has zero effect on later outcomes. In several cases, I was able to reject this null hypotheses with some confidence; in most cases, there was not enough statistical evidence for its rejection—but note that *failing to reject* the null hypothesis is not the same as *accepting it*. In these latter cases, there was also not enough statistical evidence to reject the *alternative* to the null hypothesis: that Head Start has some long-term benefit. That is, the negative findings reported here should not be taken to mean that Head Start has no long-term benefits. These findings simply mean that the statistical evidence of this study is not strong enough to reject the null hypothesis.

Among the more affirmative results, I have found that participation in Head Start increases females' educational attainment and their academic performance during the high school years. There is also weak evidence suggesting that Head Start reduces subjects' crime rates, particularly among females. Head Start appears also to improve Adult APL Survey scores among African American subjects. For other long-term outcome measures, the results are ambiguous. There may be numerous reasons for these negative findings. I personally think that major reasons are (1) the noncomparability of subjects between the HS and the No–HS groups and (2) the inadequate and unreliable measurement of family background.

The key coefficients in our regression analyses are those of *Head Start status*, *sex*, and the interaction between *Head Start status* and *sex* (*race* in the case of *Adult APL survey scores*). Let us ignore sampling error in the estimated coefficients for the time being and use the point estimates to extrapolate the magnitudes of the potential effects of Head Start, separately by subgroup (*sex* or *race*).[8] In particular, it is of interest to observe how the comparison of the HS subjects and the No–HS subjects changes with multivariate controls. That is, it is fruitful to compare unadjusted means of outcome variables to

---

[8] For statistical tests, one should go back to the regression results.

Table D.14: Unadjusted and Adjusted Means of Outcome Variables by Head Start Status and Sex

		Unadjusted Mean		Adjusted Mean	
Site/Outcome Variable	**Sex/Race**	**No-HS**	**HS**	**No-HS**	**HS**
Florida					
High school completion (%)	Male	89.3	79.0	89.3	78.3
	Female	82.7	84.8	81.1	95.1
Attainment of some postsecondary (%)	Male	47.3	33.3	47.3	36.7
	Female	51.9	46.8	51.7	64.3
Incidence of EMI (%)	Male	8.4	6.2	8.4	3.9
	Female	2.3	6.3	1.6	3.4
GPA in grades 1–4	Male	2.53	2.58	2.53	2.51
	Female	2.76	2.92	2.82	2.77
GPA in grades 5–8	Male	2.18	2.08	2.18	2.08
	Female	2.34	2.44	2.40	2.46
GPA in grades 9–12	Male	1.98	1.83	1.98	1.80
	Female	1.98	2.20	2.05	2.18
Adult APL Survey score	White	31.1	25.6	29.8	26.3
	Non-White	28.4	26.8	27.9	27.8
Arrest incidence (%)	Male	38.2	45.7	38.2	44.8
	Female	18.0	10.1	14.9	5.0
Conviction incidence (%)	Male	27.5	28.4	27.5	30.7
	Female	12.0	7.6	9.6	6.4
Receipt of public assistance (%)	Male	16.0	27.2	16.0	27.8
	Female	48.9	59.5	56.3	66.0
Colorado					
High school completion (%)	Male	88.6	70.8	88.6	84.2
	Female	84.8	75.9	79.6	83.9
Attainment of some postsecondary (%)	Male	48.6	20.6	48.6	31.6
	Female	60.6	34.5	62.7	42.1
Incidence of EMI (%)	Male	5.7	2.6	5.7	2.2
	Female	3.0	6.9	2.9	6.4
GPA in grades 1–4	Male	2.82	2.65	2.82	2.66
	Female	2.83	2.64	2.75	2.64
GPA in grades 5–8	Male	2.64	2.24	2.64	2.29
	Female	2.74	2.23	2.64	2.27
GPA in grades 9–12	Male	2.30	1.57	2.30	1.76
	Female	2.29	1.88	2.35	1.94
Adult APL Survey score	White	30.8	27.3	30.2	26.5
	Non-White	28.2	23.9	27.5	24.4
Arrest incidence (%)	Male	65.7	72.2	65.7	60.4
	Female	39.4	34.5	35.7	21.8
Conviction incidence (%)	Male	48.6	61.1	48.6	49.9
	Female	24.2	27.6	23.1	17.7
Receipt of public assistance (%)	Male	20.0	33.3	20.0	28.2
	Female	45.5	63.8	57.6	65.0

Note. Adjusted means were calculated from regression results while holding other variables included in the models at sample means. In Florida, "non–White" was African American, and in Colorado, "non–White" was Hispanic American; in both states, "White" was "White/other."

hypothetical means adjusted for the covariates included in the regression Model 3. For such a purpose, it is necessary to have a reference point at which adjusted means are calculated. For convenience, I use male, No–HS subjects as the reference group. I set the adjusted means to unadjusted (observed) means for male, No–HS subjects and then apply the estimated coefficients reported in Tables D.4–D.13 to derive adjusted means. Results from this exercise are reported in Table D.14.

Several observations are in order. First, the adjusted comparison in general makes the HS subjects look more favorable than does the unadjusted comparison. This is to be expected, for the HS subjects tended to come from families of lower SES than did the No–HS subjects. Second, there are only a few cases where the HS subjects clearly surpassed their No–HS counterparts in the adjusted mean, and these cases mostly involve females instead of males (such cases include rates of *high school completion, attainment of postsecondary education,* and *arrest incidence* among females in Florida). Third, regardless of statistical significance, the size of these multivariate effects is rather small relative to the results from the High/Scope Perry Preschool study (Schweinhart et al., 1993).

It should be emphasized again that there is a fundamental difficulty in evaluating the long-term effects of Head Start using observational data: The researcher observes only the realized outcomes of program participants, not their counterfactual outcomes of nonparticipation, and only the realized outcomes of nonparticipants, not their counterfactual outcomes of participation. As mentioned earlier, a major reason for the relatively weak results reported here is that the HS subjects and the No–HS subjects may differ in other important dimensions that are unknown to us. That is, selectivity of the two groups could seriously bias the results. Without additional information, it is hard to pin down the magnitude of the bias and assess potential damage to the results. Through the use of so-called "switching regression models," I have explicitly modeled the selectivity process with strong parametric assumptions. This exercise has yielded some intriguing results, although as a whole they are still inconclusive. Owing to the tentative nature of the exercise, I choose to describe the model and report the empirical results in an addendum (see Appendix E).

References

Manski, C. F. (1995). *Identification problems in the social sciences.* Cambridge, MA: Harvard University Press.

Manski, C. F., & Garfinkel, I. (1992). Introduction. In C. F. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs.* Cambridge, MA: Harvard University Press.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66,* 688–701.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P., with Barnett, W. S., & Epstein, A. S. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27* (Monographs of the High/Scope Educational Research Foundation, 10). Ypsilanti, MI: High/Scope Press.

Appendix E

Endogenous Switching Regression Models

Yu Xie
Professor of Sociology
University of Michigan

SELECTION ON OBSERVABLES VERSUS SELECTION ON UNOBSERVABLES

As stated in Appendix D, the major difficulty in this study is the lack of assurance that the No–HS group is comparable to the HS group. The noncomparability can lead to distortions in the estimated effects of Head Start reported in Tables D.4–D.13 in Appendix D; this is commonly referred to as selection bias. Here, I wish to make a distinction between selection on observables and selection on unobservables. Selection on observables means that the No–HS group and the HS group are different in background characteristics that are measured in the study. For example, Table D.1 on p. 140 shows that the HS subjects are more likely to be minority than are the No–HS subjects and that, compared with parents of the No–HS subjects, parents of the HS subjects tend to have less education and are more likely to use welfare assistance. These systematic differences between the two groups are controlled in the multivariate analysis reported in Tables D.4–D.13.

As discussed earlier, the No–HS subjects are not truly comparable with the HS subjects in measured characteristics: The HS subjects tended to come from families of lower socioeconomic backgrounds than did the No–HS subjects. This is a result of the original study design, because the No–HS subjects were selected from those who, as children, attended the same elementary schools as the HS subjects but did not necessarily come from families targeted for Head Start programs. It is worth emphasizing again that failure to control for such observed characteristics is likely to cause downward (i.e., negative) biases in simple estimates of the HS effects, as most observed SES characteristics are positively correlated with achievement and negatively correlated with problematic behavior and welfare dependency. This explains why the gaps between the HS group and the No–HS group in favor of the No–HS group are usually narrowed or even reversed after I control for observed covariates measuring SES, as shown in Table D.14 on p. 166.

However, it is still quite probable that the HS group and the No–HS group are not comparable on some characteristics that are unmeasured in this study. This problem is one of "selection on unobservables." There can be two explanations for the presence of selection on unobservables: (1) The measures of SES and family background (including residence) in the data set are crude and limited; for example, we do not know family income. The selection on unobserved SES characteristics is likely to induce negative biases. (2) There is the possibility that parents who enrolled their children in Head Start may have, at the outset, cared more for their children than did comparably poor parents who did not enroll their children in Head Start. Sending children to Head Start programs was a conscious decision on the part of the parents, and this decision may have been based on anticipation of beneficial outcomes for their children. This mechanism is likely to introduce positive biases and thus exaggerate the true beneficial effects of Head Start. Both of these explanations, hence, render participation in Head Start endogenous, although they affect the estimates reported in the tables of Appendix D in opposite ways. Of course, endogeneity could also come about if Head Start staff selectively recruited children of certain characteristics unknown to us. However, there is little empirical support for this conjecture.

Given the two very different explanations that could account for the problem of selection on unobservables, its consequences are unclear. If unmeasured SES characteristics dominate the selection, the estimated effects of Head Start are conservative. If parents' rational behavior dominates the selection, the estimated effects of Head Start are exaggerated. Empirically, there is no information that allows me to disentangle the two processes. Thus, I resort to data analysis to see if selection on unobservables is empirically significant for the outcomes I have examined. I will present first the model and then the empirical results.

THE SWITCHING REGRESSION MODEL

To explicitly model the endogenous process of participation in Head Start, I employ the endogenous switching regression model. Denote persons by i ($i = 1, \ldots, n$). Let z_i be a dichotomous variable that equals 1 for participation in Head Start and 0 for nonparticipation. For the ith person, let Y_i denote the outcome and X_{ki} denote the value on the kth measured independent variable ($k = 1, \ldots, K$). Assume that the X_k's are predetermined with respect to both the participation in Head Start and the outcome given participation status. Let $X_{1i} = 1$ so that the model includes an intercept term.

Now let us first model the process of Head Start participation. It is assumed that a subject has a latent tendency Z_i^* that indexes his/her likelihood of participation. With an arbitrary threshold 0, we have a measurement model that the probability of participation is $P(z_i = 1) = P(Z_i^* > 0)$. Let us now write out the reduced-form equation for the latent Z_i^* as a linear function of X_k's:

$$Z_i^* = \sum_k \pi_k X_{ki} + \varepsilon_{1i} \tag{1}$$

Equation 1 is in reduced form because both the systematic part ($\Sigma \pi_k X_{ki}$) and the stochastic part (ε_{1i}) contain structural components that can be attributed to the outcome variable (Mare & Winship, 1988). As in Appendix D, we let Y_1 and Y_0 denote outcomes, respectively, for HS participation and nonparticipation. Note that the two hypothetical outcomes pertain to the same individual. The interest centers on the expected difference between the two, namely $E(Y_1) - E(Y_0)$. We now write out two separate equations:

$$Y_{0i} = \sum_k \beta_{0k} X_{ki} + \varepsilon_{2i} \tag{2}$$

$$Y_{1i} = \sum_k \beta_{1k} X_{ki} + \varepsilon_{3i} \tag{3}$$

In these equations, the ε_{2i} and ε_{3i} denote stochastic disturbances, and the β_{0k} and β_{1k} denote parameters to be estimated. Although we hypothesize two outcomes for each individual, we of course observe only one outcome, since the other outcome is purely counterfactual. However, as argued in Appendix D, the objective of studying the treatment effect calls for the knowledge of the counterfactual outcome of HS participation for a subject who *was not* in Head Start and the counterfactual outcome of nonparticipation for someone who *was* in Head Start. Given the intrinsic unmeasurability of this knowledge, I model it statistically. Comparison of the β coefficients across equations 2 and 3 yields the treatment effect of Head Start conditional on covariates. As a special case, when $\beta_{1k} = \beta_{0k}$ for all *k*'s except for the intercept, the treatment effect is simply $(\beta_{11} - \beta_{01})$. In general, however, $\beta_{1k} \neq \beta_{0k}$. In this case, there is no single program effect.

For estimation, it is necessary to specify the structure of the reduced-form disturbances. Let us assume that ε_1, ε_2, and ε_3 follow a trivariate normal distribution with $\mathrm{Var}(\varepsilon_1) = \sigma_1^2$, $\mathrm{Var}(\varepsilon_2) = \sigma_2^2$, $\mathrm{Var}(\varepsilon_3) = \sigma_3^2$, $\mathrm{Cov}(\varepsilon_1, \varepsilon_2) = \sigma_{12}$, $\mathrm{Cov}(\varepsilon_1, \varepsilon_3) = \sigma_{13}$, and $\mathrm{Cov}(\varepsilon_2, \varepsilon_3) = \sigma_{23}$. Since ε_1 is normally distributed, equation 1 is a standard probit model. Note again that all the above parameters and equations refer to the whole population. Among the above parameters, however, all except σ_{23} and σ_1 can be estimated from observed data. For practical convenience, I normalize σ_{23} to 0 and σ_1 to 1. Given the normality assumption and the normalization constraints, the switching regression model can be estimated via maximum-likelihood (ML).

The above model described by equations 1 through 3 is called the switching regression model. Although the model is theoretically identifiable when the same independent variables (X_k's) are in all the three equations, the identifying information is very weak, and the estimates are generally unstable. To improve identifiability, it is useful to exclude a subset of the independent variables from some of the equations. For example, some variables may affect the selection equation (1) but not the substantive equations (2 and 3). In this study, I use three continuously measured contextual variables, at the U.S. Bureau of the Census tract level, as instrumental variables that affect only the selection equation, not the substantive equations.

With the switching regression model approach, selectivity can readily be seen in the estimates for the covariance terms. If there is a positive selection for Head Start participation, we expect the covariance term σ_{12} to be positive, and vice versa. If there is a positive selection *out of* Head Start, we expect σ_{12} to be negative. In other words, when $\sigma_{13} > 0$ and $\sigma_{12} < 0$, the subjects who actually enrolled in Head Start benefited more from Head Start than did the population on average, and the subjects who did not enroll in Head Start benefited more from nonparticipation than did the population on average. This would be a serious case of self-selection. If $\sigma_{13} < 0$ and $\sigma_{12} > 0$, the selection works in exactly the opposite ways. In the other two cases ($\sigma_{13} < 0$ and $\sigma_{12} < 0$, and $\sigma_{13} > 0$ and $\sigma_{12} > 0$), selectivity is not regime-specific, and selection bias is less serious (since selectivity bias for one regime (Y_0 or Y_1) cancels that for another regime (Y_1 or Y_0). Thus, estimation of the covariance terms is key to interpreting the switching regression model.

EMPIRICAL RESULTS

The switching regression model as specified in equations 1 through 3 requires that the outcome variable be continuous. For the empirical results presented in this section, I apply the switching regression model to five continuous dependent variables: *years of schooling, GPA in grades 1–4, GPA in grades 5–8, GPA in grades 9–12*, and *Adult APL Survey score*. The independent variables are similar to those used in logit and OLS models reported in Tables D.4 through D.13 of Appendix D, with one notable exception: Whereas mean variation across census tracts was controlled using dummy variables in the logit and OLS models, information at the U.S. Census tract level is used here as instrumental variables affecting only equation 1. I used three contextual variables computed from the U.S. Census: the percentage of persons living in poverty (*percent poor*), the percentage of female-headed households (*percent FHH*), and the percentage of young adults who dropped out of high school (*percent dropout*). These three variables are assumed to affect the participation probability (equation 1) but not directly affect the substantive equations (equations 2 and 3). The models were estimated using the computer software HotzTran (Avery & Hotz, 1985). I restrict the analysis to the subsample from Florida as an illustration.

Table E.1 presents the switching regression results for *years of schooling*. Equation 1 in this table, which is a probit model, predicts the probability that a subject participated in Head Start. As expected, *father's education* has a negative effect on HS participation, and *family's welfare use* has a positive effect on HS participation. However, *race, sex*, and *mother's education* have no significant effects on the participation probability. What is surprising is that not only is the effect of *mother's education* insignificant, but the point estimates are in the wrong direction (positive instead of negative). The three contextual variables are all significant, with the effects of *percent poor* and *percent dropout* being negative and the effect of *percent FHH* being positive. Note that these results should be interpreted in a multivariate framework, controlling for other factors included in the model. Equation 2 is the substantive equation for the case that a child did not participate in Head Start. Similarly, Equation 3 is the substantive equation for the case that a child did participate in Head Start.

Table E.1: Switching Regression for Years of Schooling

Variable	Equation 1 Selection Into HS (Probit Model) Coefficient	 *SE*	Equation 2 Years of Schooling No-Head Start Coefficient	 *SE*	Equation 3 Years of Schooling Head Start Coefficient	 *SE*
Constant	0.302	(0.185)	12.084	(0.184)	11.306	(0.269)
Race (1 = non-White)	0.051	(0.116)	0.398	(0.134)	0.548	(0.202)
Sex (1 = female)	0.025	(0.093)	0.018	(0.105)	0.375	(0.167)
Percent poor in tract	0.018	(0.008)				
Percent FHH in tract	0.020	(0.008)				
Percent dropout in tract	0.026	(0.007)				
Mother's education (excluded = less than high school)						
High school	0.188	(0.122)	0.265	(0.139)	0.562	(0.219)
Some postsecondary	0.108	(0.165)	0.916	(0.186)	0.385	(0.294)
Unknown	0.332	(0.151)	-0.265	(0.177)	0.324	(0.271)
Father's education (excluded = less than high school)						
High school	0.361	(0.134)	0.084	(0.162)	0.536	(0.234)
Some postsecondary	0.469	(0.192)	0.295	(0.216)	1.089	(0.356)
Unknown	0.253	(0.134)	-0.072	(0.165)	-0.203	(0.232)
Family's welfare use (excluded = never/occasional)						
Frequent	0.213	(0.106)	0.667	(0.132)	0.984	(0.176)
Unknown	0.149	(0.314)	0.238	(0.327)	0.104	(0.610)
Sigma			1.183	(0.036)	1.447	(0.057)
Rho (1, 2)/(1, 3)			0.025	(0.139)	0.013	(0.114)
***R*-squared**			0.190		0.212	
N	424		264		160	
Log-likelihood for model	1932					

The estimates for equations 2 and 3 generally agree with earlier results on *completion of high school* (Table D.4 on p. 150) and *attainment of some postsecondary education* (Table D.5 on p. 153). For example, the intercept is higher for equation 2 than for equation 3, but the effect of *sex* is positively significant for equation 3 though insignificant for equation 2. The most important estimates in Table E.1 lie in the row that is labeled "Rho (1,2)/(1,3)." Here ρ_{12} or ρ_{13} is the coefficient for the correlation in the disturbance term between equation 1 and equation 2 or between equation 1 and equation 3. It is the standardized form for σ_{12} or σ_{13} discussed earlier, measuring either positive or negative selection between a substantive equation and the selection equation. As shown in Table E.1, the two correlation coefficients are estimated to be insignificantly different from zero. Given the results, I conclude that, with respect to years of schooling, the switching regression fails to uncover any selectivity. This could happen either because there was very little selectivity to begin with or because the two sources of selectivity mentioned earlier offset each other.

Table E.2: Switching Regression for GPA in Grades 1–4

Variable	Equation 1 Selection Into HS (Probit Model) Coefficient	SE	Equation 2 GPA in Grades 1–4 No–Head Start Coefficient	SE	Equation 3 GPA in Grades 1–4 Head Start Coefficient	SE
Constant	0.096	(0.199)	2.951	(0.079)	2.770	(0.117)
Race (1 = non–White)	0.051	(0.128)	-0.581	(0.058)	0.118	(0.088)
Sex (1 = female)	0.015	(0.102)	0.261	(0.047)	0.381	(0.072)
Percent poor in tract	-0.022	(0.009)				
Percent FHH in tract	0.029	(0.009)				
Percent dropout in tract	-0.029	(0.008)				
Mother's education (excluded = less than high school)						
High school	0.195	(0.131)	0.110	(0.059)	0.219	(0.092)
Some postsecondary	0.109	(0.182)	0.331	(0.084)	0.256	(0.125)
Unknown	0.349	(0.173)	0.012	(0.082)	0.271	(0.121)
Father's education (excluded = less than high school)						
High school	-0.271	(0.143)	-0.112	(0.068)	-0.155	(0.100)
Some postsecondary	-0.271	(0.212)	-0.032	(0.096)	-0.213	(0.151)
Unknown	-0.084	(0.146)	0.154	(0.072)	-0.289	(0.099)
Family's welfare use (excluded = never/occasional)						
Frequent	0.222	(0.116)	0.022	(0.060)	-0.201	(0.175)
Unknown	-0.354	(0.367)	0.000	(0.143)	0.220	(0.297)
Sigma			0.478	(0.016)	0.566	(0.025)
Rho (1, 2)/(1, 3)			-0.067	(0.144)	-0.224	(0.130)
***R*-squared**			0.270		0.144	
N	350		216		134	
Log-likelihood for model	-949.9					

The switching regression analysis is repeated for the three measures of *GPA* as dependent variables, and the results are reported in Tables E.2 through E.4. It is interesting to note that the ρ coefficients are estimated to be negative. Among them, ρ_{13} for *GPA in grades 1–4* is sizable and statistically significant (–0.224 with a standard error of 0.130, $p < 0.10$). Although statistically insignificant, ρ_{13} for *GPA in grades 5–8* and ρ_{12} for *GPA in grades 9–12* are noticeably large (respectively, –0.125 and –0.213). These results suggest that there is some negative selection into Head Start or positive selection out of Head Start, meaning that the children who actually participated in Head Start were academically less able than the children who did not participate in Head Start. However, the evidence is weak, given the limited sample size (and hence large standard errors).

Finally, the results for *Adult APL Survey score* are presented in Table E.5, p. 177. It is apparent that the race effect differs sharply between the No–HS equation (equation 2) and the HS equation (equation 3), confirming an earlier result from Table D.10 on p. 160 that non–White (African

Table E.3: Switching Regression for GPA in Grades 5–8

	Equation 1 Selection Into HS (Probit Model)		Equation 2 GPA in Grades 5–8 No–Head Start		Equation 3 GPA in Grades 5–8 Head Start	
Variable	**Coefficient**	***SE***	**Coefficient**	***SE***	**Coefficient**	***SE***
Constant	0.125	(0.198)	2.539	(0.090)	2.235	(0.135)
Race (1 = non-White)	0.024	(0.127)	-0.594	(0.067)	-0.140	(0.102)
Sex (1 = female)	0.042	(0.101)	0.212	(0.054)	0.395	(0.084)
Percent poor in tract	-0.022	(0.008)				
Percent FHH in tract	0.029	(0.009)				
Percent dropout in tract	-0.026	(0.008)				
Mother's education (excluded = less than high school)						
High school	0.233	(0.131)	0.190	(0.068)	0.091	(0.108)
Some postsecondary	0.082	(0.180)	0.270	(0.094)	0.169	(0.147)
Unknown	0.386	(0.170)	0.015	(0.093)	0.297	(0.139)
Father's education (excluded = less than high school)						
High school	-0.410	(0.143)	-0.057	(0.079)	0.077	(0.118)
Some postsecondary	-0.374	(0.209)	0.142	(0.110)	0.280	(0.176)
Unknown	-0.203	(0.144)	-0.042	(0.082)	-0.115	(0.113)
Family's welfare use (excluded = never/occasional)						
Frequent	0.192	(0.116)	-0.124	(0.068)	-0.441	(0.082)
Unknown	-0.291	(0.379)	-0.290	(0.179)	0.157	(0.348)
Sigma			0.549	(0.019)	0.665	(0.029)
Rho (1, 2)/(1, 3)			-0.096	(0.143)	-0.125	(0.123)
***R*-squared**			0.243		0.186	
N	355		220		135	
Log-likelihood for model	-1067.4					

American) participants benefit more from Head Start than do White/other participants. The correlation coefficient between the selection equation and the second equation is estimated at 0.144, although it is not statistically significant at the 0.10 level.

CONCLUSION

The main difficulty of this study lies in the noncomparability of the HS group and the No–HS group. This problem of noncomparability can be further decomposed into two forms: selection on observables and selection on unobservables. The multivariate framework with logit and OLS regressions presented in Tables D.4–D.13 in Appendix D aids in reducing biases associated with selection on observables, although there are reasons to assume that substantial biases remain after the control of the observed variables. How to control for selection on unobservables is less straightforward, largely because we do not have any useful knowledge about unobservable variables. Hence,

Table E.4: Switching Regression for GPA in Grades 9–12

	Equation 1 Selection Into HS (Probit Model)		Equation 2 GPA in Grades 9–12 No–Head Start		Equation 3 GPA in Grades 9–12 Head Start	
Variable	**Coefficient**	***SE***	**Coefficient**	***SE***	**Coefficient**	***SE***
Constant	0.124	(0.198)	2.169	(0.113)	2.045	(0.159)
Race (1 = non-White)	0.099	(0.129)	-0.374	(0.085)	-0.357	(0.123)
Sex (1 = female)	0.005	(0.102)	0.059	(0.069)	0.384	(0.097)
Percent poor in tract	-0.022	(0.009)				
Percent FHH in tract	0.028	(0.009)				
Percent dropout in tract	-0.031	(0.008)				
Mother's education (excluded = less than high school)						
High school	0.222	(0.133)	0.308	(0.088)	0.030	(0.127)
Some postsecondary	0.135	(0.178)	0.312	(0.121)	-0.012	(0.164)
Unknown	0.335	(0.174)	0.169	(0.119)	0.027	(0.164)
Father's education (excluded = less than high school)						
High school	-0.261	(0.144)	-0.115	(0.100)	0.317	(0.135)
Some postsecondary	-0.285	(0.208)	0.019	(0.140)	0.617	(0.195)
Unknown	-0.140	(0.146)	-0.187	(0.104)	0.149	(0.136)
Family's welfare use (excluded = never/occasional)						
Frequent	0.066	(0.121)	-0.223	(0.086)	-0.440	(0.107)
Unknown	-0.330	(0.364)	0.334	(0.211)	0.442	(0.395)
Sigma			0.709	(0.024)	0.754	(0.033)
Rho (1, 2)/(1, 3)			-0.213	(0.145)	-0.092	(0.122)
***R*-squared**			0.123		0.183	
N	352		223		129	
Log-likelihood for model	-1197.2					

strong parametric assumptions need to be brought in to account for biases associated with selection on unobservables (for a recent review, see Winship & Mare, 1992).

In this appendix, I introduce the switching regression as an explicit modeling design to test selectivity between participation in Head Start and later outcomes. I present the switching regression results for five continuous variables: *years of schooling, GPA in grades 1–4, GPA in grades 5–8, GPA in grades 9–12,* and *Adult APL Survey score.* The analysis is restricted to the subsample from Florida, since the Colorado subsample is too small for the results to be statistically significant. Even with the Florida subsample, the results are statistically too imprecise to be conclusive.

The models with the GPA measures as dependent variables come closest to revealing selectivity in the data. This is especially true if we relax the conventional criteria for statistical significance. The point estimates from the models for the three GPA measures suggest that there exists some selectivity

Table E.5: Switching Regression for Adult APL Survey Score

	Equation 1 Selection Into HS (Probit Model)		Equation 2 APL Score No–Head Start		Equation 3 APL Score Head Start	
Variable	**Coefficient**	***SE***	**Coefficient**	***SE***	**Coefficient**	***SE***
Constant	0.295	0.191	27.347	0.942	26.557	1.204
Race (1 = non–White)	-0.064	0.119	-2.573	0.679	2.183	0.895
Sex (1 = female)	0.084	0.096	2.921	0.536	2.754	0.749
Percent poor in tract	-0.020	0.008				
Percent FHH in tract	0.022	0.008				
Percent dropout in tract	-0.027	0.008				
Mother's education (excluded = less than high school)						
High school	0.162	0.127	1.824	0.700	1.448	0.996
Some postsecondary	0.152	0.171	1.809	0.949	2.402	1.320
Unknown	0.459	0.157	-0.008	0.916	-3.958	1.223
Father's education (excluded = less than high school)						
High school	-0.356	0.139	2.398	0.823	-2.096	1.060
Some postsecondary	-0.447	0.196	3.224	1.093	-0.230	1.564
Unknown	-0.304	0.139	0.888	0.838	-5.279	1.056
Family's welfare use (excluded = never/occasional)						
Frequent	0.258	0.110	-1.125	0.678	-1.789	0.791
Unknown	-0.425	0.355	5.282	1.631	-1.403	3.279
Sigma			5.899	0.189	6.307	0.258
Rho (1, 2)/(1, 3)			0.144	0.154	0.006	0.119
R-squared			0.161		0.311	
N	395		246		149	
Log-likelihood for model	-3022.5					

in the direction of favoring children who did not participate in Head Start. Across the models, either (1) the correlation in disturbance between the participation equation and the HS equation for GPA or (2) the correlation in disturbance between the participation equation and the No–HS equation for GPA is estimated negative. A negative estimate of the former indicates *negative* selection into HS participation; a negative estimate of the latter indicates *positive* selection into nonparticipation. However, given these estimates' rather large standard errors (*SE* is larger than half of the point estimate in every case), I do not wish to draw a firm conclusion from this exercise. In sum, there appears some evidence in favor of the hypothesis that HS children may be more disadvantaged than No–HS children in unmeasured but relevant characteristics. However, because the evidence is statistically weak, I do not wish to put too much faith in these results and therefore present them only in this appendix as an addendum to my report *Assessment of the Long-Term Benefits of Head Start* (Appendix D).

References

Avery, R. B., & Hotz, V. J. (1985). *HotzTran: User's manual.* Old Greenwich, CT: CERA Economic Consultants, Inc.

Mare, R. D., & Winship, C. (1988). Endogenous switching regression models for the causes and effects of discrete variables. In J. S. Long (Ed.), *Common problems/proper solutions: Avoiding error in quantitative research* (pp. 132–160). Newbury Park, CA: Sage.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology, 18,* 327–350.

Appendix F

Estimating the Long-Term Effects of Head Start

Sue M. Marcus
Assistant Professor of Biostatistics
University of Pennsylvania School of Medicine

It has long been recognized that evaluations of compensatory education programs tend to result in underestimation of effects (Campbell & Boruch, 1975). In particular, this was a concern for the Long-Term Benefits of Head Start (LTBHS) study, since it appeared that the most severely impoverished children had been recruited for Head Start, thus leaving a pool of potential comparison subjects who were relatively less disadvantaged. Consequently, the estimate of the benefits of Head Start could be misleading, since the ostensible effects of Head Start may be at least in part a reflection of study group differences. Thus, the technical advisory panel for the LTBHS study asked me to conduct a special analysis using propensity score analysis to address this problem of potential selection bias.

This appendix presents the results of my analysis. The focus is on the problem of underestimation of effects due to selection bias. The analysis looks for evidence of selection bias. Estimates of the long-term effects of Head Start are adjusted with respect to bias due to measured covariates. In addition, I consider the uncertainty in assessing the long-term effects of Head Start due to hidden bias resulting from unobserved differences between the Head Start subjects and the controls.

METHODOLOGICAL CONSIDERATIONS

An observational study is a comparison of a treatment group and a control group that is designed to evaluate the effect of the treatment when it is not feasible to randomly assign subjects to treatment or control. Several observational studies of primary interest are nested within the LTBHS study. In each of the Colorado and Florida studies, those who attended Head Start can be considered as the treatment group, and those who attended no preschool program can be thought of as the control (No–Head Start) group. Additionally, in the Florida study, those who attended Head Start programs using the High/Scope educational approach (Model Head Start) can be considered a treatment group, while those who attended other Head Start programs (Regular Head Start) can be considered a control group.

Cochran (1965) presented methodological issues that arise in observational studies. The basic problem deals with possible noncomparability of the treatment and control groups before the treatment. *If the treatment group and the control group are not similar at base line, group differences in outcome may reflect, at least in part, the original group differences as well as the effect of the treatment.* In randomized studies, in which subjects are randomly assigned to the treatment and control groups, the groups should be, at least on the average, similar due to the randomization (Fisher, 1935). However, in observational studies there is no process to ensure similarity of the study groups.

In the LTBHS study, there was no randomization to Head Start versus No–Head Start (or to Model Head Start versus Regular Head Start). Instead, the study groups were composed of young adults who, in 1971, lived in the same neighborhoods and entered the same elementary schools. The groups were determined by participation in Regular Head Start programs in 1970–71, participation in Model Head Start programs in 1970–71, or no participation in Head Start or any other preschool program in 1970–71 or any other year. Furthermore, the Head Start staff from 1970 reported that they conducted the recruitment for Head Start in the most impoverished neighborhoods, and therefore it is likely that a bias might result from the lower socioeconomic status of those in the Head Start group (see Oden, 1995). Thus, the comparison of the treatment groups at base line, which follows in the next section, is crucial.

COMPARISON OF TREATMENT GROUPS AT BASE LINE

The advisory panel proposed three comparisons of interest: a Head Start group versus a No–Head Start group in Florida, a Head Start group versus a No–Head Start group in Colorado, and a Model Head Start group versus a Regular Head Start group in Florida. The study groups were compared with respect to *gender; race; household composition;* and *parents' education, job status,* and *welfare use.* Group differences were tested using t-tests and χ^2 tests, as appropriate. Tables F.1 and F.2 display these results.

In Florida, compared with the No–Head Start group, the Head Start group tended to have more children but fewer adults in each household. Furthermore, compared with No–Head Start subjects, the Head Start subjects had a significantly higher proportion of fathers who dropped out of high school (44% vs. 29%, $p < .01$). The mean number of years of education was significantly lower for fathers of subjects who had attended Head Start than for fathers of No–Head Start subjects (10.4 vs. 11.5, $p < .001$). In addition, the Head Start group had a significantly lower proportion of parents who had never been on welfare than did the No–Head Start group (46% vs. 59%, $p < .01$). Thus, it appears that members of the Florida Head Start group came from more disadvantaged families than did the Florida control subjects.

Also, in Colorado, compared with the No–Head Start subjects, the Head Start subjects appear to have come from families with more poverty and less education. The Head Start group was about three-quarters Hispanic

Table F.1: Comparison of Treatment and Control Groups (Dichotomous Covariates)

		Florida		Colorado	
Covariate	**Head Start Status**	***N***	**Proportion**	***N***	**Proportion**
Gender (male)	Head Start	160	.51	130	.55
	No-Head Start	264	.50	68	.51
	Model Head Start	74	.45		
	Regular Head Start	86	.56		
Non-White[a]	Head Start	160	.78	130	.73***
	No-Head Start	264	.80	68	.32
	Model Head Start	74	.73		
	Regular Head Start	86	.83		
Two-parent household	Head Start	160	.64	130	.79
	No-Head Start	264	.67	68	.79
	Model Head Start	74	.64		
	Regular Head Start	86	.64		
Mother: High school dropout	Head Start	126	.34	114	.64***
	No-Head Start	221	.32	58	.21
	Model Head Start	52	.40		
	Regular Head Start	74	.30		
Mother: Some postsecondary education	Head Start	126	.16	114	.12*
	No-Head Start	221	.19	58	.24
	Model Head Start	52	.12		
	Regular Head Start	74	.19		
Father: High school dropout	Head Start	99	.44**	93	.65**
	No-Head Start	167	.29	58	.40
	Model Head Start	46	.52		
	Regular Head Start	53	.38		
Father: Some postsecondary education	Head Start	99	.12	93	.08**
	No-Head Start	167	.19	58	.19
	Model Head Start	46	.11		
	Regular Head Start	53	.13		
Parents never on welfare	Head Start	160	.46**	130	.48**
	No-Head Start	264	.59	68	.66
	Model Head Start	74	.43		
	Regular Head Start	86	.48		
Parents on welfare frequently	Head Start	160	.33+	130	.24
	No-Head Start	264	.24	68	.15
	Model Head Start	74	.32		
	Regular Head Start	86	.33		

[a] In Florida, the non-White subjects were African American; in Colorado, they were Hispanic American.

+*p* < .10. **p* < .05. ***p* < .01. ****p* < .001.

Table F.2: Comparison of Treatment and Control Groups (Nondichotomous Covariates)

		Florida			Colorado		
Covariate	**Head Start Status**	***N***	**Mean**	***SD***	***N***	**Mean**	***SD***
Age at time of interview	Head Start	160	22.2	1.5	130	22.1	1.0
	No-Head Start	264	22.2	1.3	68	22.1	1.1
	Model Head Start	74	22.0	1.7			
	Regular Head Start	86	22.3	1.3			
Number of siblings	Head Start	153	4.2*	2.7	130	3.9***	2.5
	No-Head Start	261	3.6	2.6	68	2.6	1.8
	Model Head Start	72	4.4	2.7			
	Regular Head Start	81	4.0	2.6			
Number of children in household	Head Start	160	4.5**	2.5	130	3.8***	2.2
	No-Head Start	264	3.9	2.5	68	2.8	1.5
	Model Head Start	74	4.8	2.6			
	Regular Head Start	86	4.3	2.4			
Number of adults in household	Head Start	160	2.0*	0.9	130	2.0	1.3
	No-Head Start	264	2.2	1.5	68	2.1	0.8
	Model Head Start	74	2.0	0.9			
	Regular Head Start	86	2.0	0.9			
Mother's job when subject was 4–6	Head Start	85	21.2	17.3	56	23.4+	18.7
	No-Head Start	124	24.5	17.0	26	31.4	22.8
	Model Head Start	37	21.5	16.9			
	Regular Head Start	48	21.0	17.8			
Father's job when subject was 4–6	Head Start	99	19.2	16.0	104	15.2**	10.9
	No-Head Start	175	21.7	16.1	58	22.2	18.9
	Model Head Start	44	20.9	17.0			
	Regular Head Start	55	17.9	15.2			
Mother's years of education	Head Start	126	11.3	2.1	114	9.6***	3.0
	No-Head Start	221	11.5	2.4	58	11.9	2.2
	Model Head Start	52	11.0	2.0			
	Regular Head Start	74	11.5	2.2			
Father's years of education	Head Start	99	10.4***	2.7	93	9.5***	3.2
	No-Head Start	167	11.5	2.5	58	11.1	2.8
	Model Head Start	46	9.9+	2.8			
	Regular Head Start	53	10.9	2.6			

+$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

American, while the control group was only one-third Hispanic American ($p < .001$). The Head Start subjects were from families who had significantly more children than the control group families had ($p < .001$). Compared with the No–Head Start group, the Head Start group had more mothers who had dropped out of high school (64% vs. 21%, $p < .001$), more fathers who had dropped out of high school (65% vs. 40%, $p < .01$), fewer mothers with some postsecondary education (12% vs. 24%, $p < .05$), and fewer fathers with some postsecondary education (8% vs. 19%, $p < .01$). The Head Start mothers had significantly fewer years of education than the No–Head Start mothers had (9.6 vs. 11.9, $p < .001$). Similarly, the Head Start fathers had significantly fewer years of education than the No–Head Start fathers had (9.5 vs. 11.1, $p < .001$). Compared with No–Head Start parents, a smaller proportion of Head Start parents had never been on welfare (48% vs. 66%, $p < .01$). Furthermore, the Head Start parents tended to have lower Duncan SEI occupational classification scores, which is additional evidence that the Head Start families were more disadvantaged.

In Florida, the Model Head Start group had more mothers who were high school dropouts than did the Regular Head Start group (40% vs. 30%), and the Model Head Start group had more fathers who were high school dropouts than did the Regular Head Start group (52% vs. 38%). These differences were not statistically significant, but the sample size for this comparison is smaller than the sample size for the comparisons of Head Start versus No–Head Start.

In summary, there is evidence of selection bias due to Head Start recruitment from the most impoverished neighborhoods. The Head Start subjects appear to have been from families who tended to be more disadvantaged than those of the No–Head Start subjects. The difference between the Head Start group and the No–Head Start group is significant for the Florida study and even more substantial for the Colorado study. There is also a hint that the Florida Model Head Start group had a lower socioeconomic level than the Florida Regular Head Start group, as evidenced by the lower educational levels of the Model Head Start parents.

These findings are important in that *all* of these biases are *against* finding positive program effects and *for* finding spurious negative program effects.

COMPARISON OF TREATMENT AND CONTROL GROUPS AS YOUNG ADULTS

The study groups were compared with respect to 12 selected outcome variables—*high school graduation, some postsecondary education, ever on welfare, employed or in school at time of interview, teen parenthood, years of education, GPA in elementary school, GPA in middle school, GPA in high school, Adult APL Survey score, number of arrests,* and *number of convictions* (see Tables F.3 and F.4). Since the treatment groups were not comparable at base line, *these tables must be interpreted with great caution;* this cannot be stressed too

Table F.3: Comparison of Treatment and Control Groups (Dichotomous Outcomes)

		Florida		Colorado	
Variable	**Head Start Status**	***N***	**Proportion**	***N***	**Proportion**
High school graduation	Head Start	160	.82	130	$.73^{*}$
	No–Head Start	264	.86	68	.87
	Model Head Start	74	.78		
	Regular Head Start	86	.85		
Some postsecondary education	Head Start	160	$.40^{*}$	130	$.27^{***}$
	No–Head Start	264	.50	68	.54
	Model Head Start	74	$.32^{+}$		
	Regular Head Start	86	.47		
Ever on welfare	Head Start	160	$.43^{*}$	130	$.47^{*}$
	No–Head Start	264	.33	68	.32
	Model Head Start	74	.46		
	Regular Head Start	86	.41		
Employed or in school at time of interview	Head Start	160	.73	129	.66
	No–Head Start	264	.68	68	.76
	Model Head Start	74	.70		
	Regular Head Start	86	.74		
Teen parenthood	Head Start	157	.27	130	.22
	No–Head Start	261	.20	68	.21
	Model Head Start	74	.26		
	Regular Head Start	83	.28		

Note. Since the study groups are not comparable at base line, these tables must be interpreted with caution.
$^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

strongly. Because the Head Start subjects tended to come from families who had relatively lower socioeconomic status, differences in outcomes will reflect these baseline differences, as well as any effects of Head Start.

For the Florida Head Start versus No–Head Start comparison, the Head Start group had a significantly lower proportion of subjects with *some postsecondary education* (40% vs. 50%, $p < .05$), significantly fewer *years of education* (12.09 vs. 12.50, $p < .001$), significantly lower *Adult APL Survey scores* (26.5 vs. 29.0, $p < .001$), and a significantly higher proportion of subjects who were *ever on welfare* (43% vs. 33%, $p < .05$).

The Colorado Head Start versus No–Head Start comparison revealed that the Head Start group tended to have a significantly lower rate of *high school graduation* (73% vs. 87%, $p < .05$), a significantly lower proportion of subjects with *some postsecondary education* (27% vs. 54%, $p < .001$), significantly fewer *years of education* (11.84 vs. 12.75, $p < .001$), significantly lower *GPAs in elementary, middle,* and *high school* years ($p < .001$), significantly lower *Adult APL Survey scores* (24.7 vs. 29.9, $p < .001$), significantly more subjects

Table F.4: Comparison of Treatment and Control Groups (Nondichotomous Outcomes)

		Florida			Colorado		
Variable	**Head Start Status**	***N***	**Mean**	***SD***	***N***	**Mean**	***SD***
Years of education	Head Start	160	12.09***	1.64	130	11.84***	1.39
	No-Head Start	264	12.50	1.32	68	12.75	1.55
	Model Head Start	74	11.92	1.77			
	Regular Head Start	86	12.24	1.51			
GPA in elementary school	Head Start	134	2.74	.62	108	2.64***	.38
	No-Head Start	216	2.64	.56	53	2.83	.39
	Model Head Start	63	3.15***	.32			
	Regular Head Start	71	2.38	.59			
GPA in middle school	Head Start	135	2.25	.74	98	2.24***	.68
	No-Head Start	220	2.26	.63	48	2.69	.73
	Model Head Start	65	2.53***	.66			
	Regular Head Start	70	2.00	.72			
GPA in high school	Head Start	129	2.01	.84	88	1.71***	.94
	No-Head Start	223	1.98	.76	40	2.29	.98
	Model Head Start	60	2.31***	.82			
	Regular Head Start	69	1.75	.77			
Adult APL Survey score	Head Start	149	26.5***	7.63	120	24.7***	8.51
	No-Head Start	246	29.0	6.45	61	29.9	7.46
	Model Head Start	69	25.70	7.78			
	Regular Head Start	80	27.21	7.47			
Number of arrests	Head Start	160	1.36	3.00	130	3.88**	6.02
	No-Head Start	264	1.25	3.46	68	2.13	3.18
	Model Head Start	74	.95+	2.30			
	Regular Head Start	86	1.72	3.46			
Number of convictions	Head Start	160	1.00	2.74	130	1.93***	3.12
	No-Head Start	264	.95	3.19	68	.84	1.60
	Model Head Start	74	.54*	1.59			
	Regular Head Start	86	1.41	3.40			

Note. Since the study groups are not comparable at base line, these tables must be interpreted with caution.
+$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

ever on welfare (47% vs. 32%, $p < .05$), and a significantly greater *number of arrests* and *convictions* (3.88 vs. 2.13, $p < .01$, and 1.93 vs. .84, $p < .001$, respectively).

Thus, without taking baseline study groups into consideration, the story looks quite grim for Head Start. In fact, participation in Head Start appears to be harmful! However, as noted earlier, the differences in outcomes are most likely due to the generally lower socioeconomic levels of the Head Start population at base line. *Consequently, any estimate of the effect of Head Start that is not adjusted for baseline group differences is dead wrong!* Clearly, a more sophisticated analysis that takes into consideration the baseline socioeconomic differences is appropriate.

The Model Head Start versus Regular Head Start comparison showed that the Model (High/Scope) group had significantly higher mean *GPAs for elementary school* (3.15 vs. 2.38, $p < .001$), *middle school* (2.53 vs. 2.00, $p < .001$), and *high school* (2.31 vs. 1.75, $p < .001$) and a significantly lower mean *number of convictions* for crime (.54 vs. 1.41, $p < .05$). Since at base line the Regular Head Start group did not appear to be at greater risk compared with the Model Head Start group, it seems reasonable that the difference in outcomes between the groups reflects actual benefits of Model Head Start, rather than baseline group differences.

METHODOLOGY FOR ESTIMATING THE EFFECTS OF HEAD START

To assess the effects of Head Start, odds ratios were estimated for the binary outcomes, and treatment effects were estimated for the nonbinary outcomes. An odds ratio for a given binary outcome gives the odds that a subject with the outcome is in Head Start; an odds ratio greater than 1 indicates that the outcome is more likely for those in Head Start. The estimated treatment effect is the treatment-group versus control-group difference in the mean value for a given outcome; a positive treatment effect indicates that the Head Start group experienced a higher mean value of the outcome than did the control group.

As stated earlier, appropriate estimates of the effects of Head Start need to take baseline group differences into consideration. Cochran (1965) proposed three ways of adjusting for treatment group differences: (1) matching, (2) subclassification, and (3) covariance adjustment. These three methods were used to estimate the odds ratios and treatment effects, adjusted for observed baseline differences, and will be described in the sections that follow.

Use of Propensity Scores

An effective way to adjust (using either matching, subclassification, or covariance adjustment) is through use of the propensity score (Rosenbaum & Rubin, 1983). A propensity score is the probability of being in the treatment group, given a set of observed covariates. Propensity scores can be estimated from the data by using, for example, logistic regression (Cox, 1970). Then matching, subclassification, or covariance adjustment on the propensity score tends to balance the treatment groups, at least with respect to the observed covariates used in estimating the propensity scores.

For this analysis, propensity scores were calculated separately for each of the three planned comparisons by using logistic regression. The covariates listed earlier were used in constructing the propensity scores; however, since many of the subjects had missing information ("unknown" responses) concerning their parents' education (182 out of 424 for the Florida Head Start vs. No–Head Start comparison, 53 out of 198 for the Colorado Head Start vs. No–Head Start comparison, and 73 out of 160 for the Florida Model Head Start vs. Regular Head Start comparison), a separate logistic regression that excluded covariates regarding the parents' education was used to calculate the propensity scores for these subjects.

Propensity scores are quite useful in this situation in that they provide a good univariate measure of socioeconomic status. Table F.5 illustrates this using the Florida Head Start versus No–Head Start comparison. Propensity scores were calculated for this comparison, and then all subjects were assigned to one of 5 quintiles on the basis of the propensity scores. The first quintile (Q1) consists of the subjects *least* similar to those in Head Start, while Q5 should be the subjects *most* similar to those in Head Start. It is as expected that those in Q1 (most like the control group) tend to have fathers that have relatively more education and tend to have families who have been relatively less often on welfare; those in Q5 (most like the Head Start subjects) tend to have fathers with less education and tend to have families who have been on welfare more frequently.

Covariance Adjustment

It has been shown (Rosenbaum & Rubin, 1983) that adjustment for the propensity score is sufficient to remove bias due to measured covariates. For this analysis, each outcome was modeled (with linear or logistic regression) as a function of a treatment group indicator (a binary variable equal to 1 for those in the treatment group, equal to 0 for those in the control group) and the propensity score. When linear regression is used, the coefficient for the

Table F.5: Florida Head Start Versus No–Head Start—Selected Outcomes by Propensity Score Quintiles

Covariate/Quintile	Head Start	No–Head Start
Father's years of schooling		
Propensity score Q1	12.8	12.8
Propensity score Q2	11.5	11.8
Propensity score Q3	11.2	11.5
Propensity score Q4	11.0	9.7
Propensity score Q5	8.9	9.1
Parents never on welfare		
Propensity score Q1	.94	.88
Propensity score Q2	.76	.86
Propensity score Q3	.62	.59
Propensity score Q4	.25	.24
Propensity score Q5	.17	.09

treatment group indicator is the estimated treatment effect, adjusted for those measured covariates used in estimating the propensity scores. If logistic regression is used, the coefficient for the treatment group indicator will yield the log-odds ratio (which can easily be transformed into the odds ratio), adjusted for the measured covariates.

Matching on the Propensity Score

Matching is a way of associating a treated individual with a control individual who is similar with respect to measured characteristics. This is a (partial) solution to the problem of study group noncomparability, since a matched analysis involves comparison groups that tend to be balanced with respect to the observed characteristics.

On the one hand, "greedy" matching essentially finds the best matches for some individuals, at the expense of good matches for others. On the other hand, optimal matching, proposed by Rosenbaum (1989), minimizes overall the differences between matched treated subjects and controls.

Commonly, 1 treated subject will be matched with 1 control subject (1-to-1 matching) or with *k* individuals (1-to-*k* matching). However a better matching procedure is full matching, in which either a treated unit may be matched to 1 or more controls, or a control unit may be matched to 1 or more treated units (Gu & Rosenbaum, 1993; Rosenbaum,1989). In studies such as the LTBHS study, where the study groups are quite different at base line, full matching should provide better comparability than 1-to-1 matching provides.

For the purposes of this analysis, optimal full matching on the propensity score was performed for each of the three comparisons—Florida Head Start versus No–Head Start, Colorado Head Start versus No–Head Start, and Florida Model Head Start versus Regular Head Start. Since each match is composed of 1 treated individual and 1 or more control individuals (or 1 control individual and 1 or more treated individuals) who are closely matched on the propensity score, the result is comparison groups that are balanced with respect to the observed measures. Thus, the problem of noncomparability of study groups is addressed, at least for the observed covariates.

Analysis by matched sets was then carried out using either a Mantel-Haenszel analysis for each binary outcome or a mixed-effect model (with the match considered as a random effect) for the other outcomes. Thus, each matched set of similar individuals is considered as a unit of the analysis; so, intuitively, the comparison groups are balanced with respect to the measured covariates.

Subclassification on the Propensity Score

Another method for controlling systematic differences between the treatment and control groups is subclassification. A subclassification determines subclasses of control and treated individuals who are comparable within each subclass with respect to observed covariates. Then treated and control units within the same subclass can be compared directly. If subclasses are formed from the propensity score, all covariates should be balanced (Rosenbaum & Rubin, 1984). For this analysis, 5 subclasses were determined by propensity score quintiles. Analysis by subclass was performed by using either the Mantel-Haenszel test for the binary outcomes or a linear model for the other outcomes.

Adjusted Effects

For the Head Start versus No–Head Start comparisons, the unadjusted effects must be viewed with caution, since the comparison groups are remarkably different in important ways. But the adjusted rates should give more appropriate estimates of the actual effects of Head Start (or Model Head Start). Tables F.6 and F.7 give the unadjusted and adjusted results for the three planned comparisons of interest.

Table F.6: Unadjusted and Adjusted Odds Ratios for Dichotomous Outcomes

Variable/Analysis	Florida Head Start vs. No–Head Start	Colorado Head Start vs. No–Head Start	Model vs. Regular Head Start
High school graduation			
Unadjusted	.74	.41*	.65
Covariance-adjusted	1.11	.96	.60
Matching	1.27	.94	.73
Subclassification	.97	.96	.55
Some postsecondary education			
Unadjusted	.68**	.31***	.55+
Covariance-adjusted	.87	.61	.51+
Matching	.80	.60	.50+
Subclassification	.82	.60	.52
Ever on welfare			
Unadjusted	1.57	1.85	1.24
Covariance-adjusted	1.29	.92	.99
Matching	1.28	.75	.91
Subclassification	1.27	.88	.92
Employed or in school at time of interview			
Unadjusted	1.23	.59	.81
Covariance-adjusted	1.47+	.67	1.16
Matching	1.60+	.68	1.16
Subclassification	1.48+	.70	1.24
Teen parenthood			
Unadjusted	1.47	1.107	.90
Covariance-adjusted	1.15	.78	1.03
Matching	1.12	.76	.86
Subclassification	1.15	.79	.84

Note. Odds ratio greater than 1 favors Head Start.
+*p* < .10. **p* < .05. ***p* < .01. ****p* < .001.

Table F.7: Unadjusted and Adjusted Treatment Effects for Continuous Outcomes

Variable/Analysis	Florida Head Start vs. No-Head Start	Colorado Head Start vs. No-Head Start	Model vs. Regular Head Start
Education in years			
Unadjusted	-.41**	-.91***	-.33
Covariance-adjusted	-.16	-.37	-.27
Matching	-.17	-.47+	-.25
Subclassification	-.22	-.20	-.34
GPA in elementary school			
Unadjusted	.10	-.18**	.77***
Covariance-adjusted	.12+	-.07	.70***
Matching	.13*	-.09	.75***
Subclassification	.10	-.03	.71***
GPA in middle school			
Unadjusted	-.01	-.45***	.53***
Covariance-adjusted	.06	-.26+	.51***
Matching	.10	-.22+	.49**
Subclassification	.04	-.46***	.38*
GPA in high school			
Unadjusted	.03	-.58***	.56***
Covariance-adjusted	.11	-.34+	.60***
Matching	.14	-.38+	.66***
Subclassification	.07	-.41+	.62***
Adult APL Survey score			
Unadjusted	-2.46***	-5.18***	-1.52
Covariance-adjusted	-1.89**	-2.96*	-2.15
Matching	-1.78*	-2.90+	-2.01
Subclassification	-1.98**	-3.14*	-4.23*
Number of arrests			
Unadjusted	.11	1.74*	-.77+
Covariance-adjusted	.11	1.01	-.96+
Matching	.04	1.05	-1.20*
Subclassification	.17	.50	-.56+
Number of convictions			
Unadjusted	.05	1.09**	-8.66*
Covariance-adjusted	.05	.75	-1.03*
Matching	.03	.82	-1.31*
Subclassification	.08	.59	-1.08*

+p < .10. *p < .05. **p < .01. ***p < .001.

For the Florida and Colorado Head Start versus No–Head Start comparisons, without adjustment for baseline group differences, it appears that Head Start tended to produce a significantly lower proprotion of subjects who went on to *some postsecondary education;* however, the ostensible effect of Head Start is most likely due to baseline group differences rather than to the effect of Head Start, since the adjusted odds ratios are not statistically significant. Similarly, without taking baseline differences into consideration, it appears that for the Colorado comparison, Head Start subjects were less likely to achieve *high school graduation;* but after adjustment, the Head Start and No–Head Start groups seem to be essentially the same with respect to *high school graduation.* In addition, many other ostensible Head Start "detriments" (e.g., for Florida, fewer *years of education,* and for Colorado, fewer *years of education,* lower *GPA in elementary school,* lower *GPA in high school,* greater *numbers of arrests* and *convictions*) probably reflect the lower socioeconomic status of the Head Start subjects rather than the effect of Head Start, since adjustment for observed baseline differences reveals no significant differences in the outcomes.

Which effects remain significant after adjustment? Florida Head Start seems significantly beneficial with respect to *GPA in elementary school* after matching adjustment. No–Head Start seems to be significantly better with respect to *Adult APL Survey score* for both Florida and Colorado.

The Florida Model versus Regular Head Start groups were similar at base line, so the adjustment makes little difference with these comparisons. Model Head Start produced significantly higher *GPAs in elementary school, middle school* and *high school;* moreover adjustment does not weaken these inferences. After adjustment, the Florida Model Head Start group also had significantly lower *numbers of arrests* and *convictions.*

In summary, after adjustment for observed covariates, there appears to be very little difference in outcomes between those who attended Head Start and those who did not, at both the Florida site and the Colorado site. However, it seems that compared with those in regular Head Start, those in Model Head Start had higher *GPAs in elementary school, middle school* and *high school.* Furthermore, the Model Head Start group had lower *numbers of arrests* and *convictions.*

HIDDEN BIAS

Rosenbaum (1995) considers an analysis of a study of cigarette smoking and mortality (Cochran, 1968) wherein Cochran removes overt bias in the estimate of mortality by adjusting for an imbalance in the covariate age. Rosenbaum's discussion goes as follows:

> Which rates should be trusted, unadjusted or adjusted? Neither. The unadjusted rates are clearly wrong as a basis for estimating the effects of smoking, for they compare men who are not comparable in terms of one of the most important features of human mortality, namely age. The adjusted rates are not

> clearly wrong. They might estimate the effects of smoking. However, it is possible that there is another covariate that was not recorded that has an impact similar to age; in this case, there would be a hidden bias. (p. 60)

Similarly, the unadjusted effects of Head Start are clearly wrong. The adjusted rates (Tables F.6 and F.7) may be good estimates of the long-term effects of Head Start. However, it is also possible that there is hidden bias that is due to factors that were not measured. For example, the Head Start children appeared to be from more impoverished and less educated homes. Perhaps they also differed from the control group in other ways that were not measured, e.g., in parents' attitudes concerning the importance of education.

If there is hidden bias that is due to an imbalance between the study groups with respect to one or more unobserved covariates, there is no way to adjust for this, since the information necessary for adjustment is unavailable. However, sensitivity analyses for observational studies (Marcus, 1997; Rosenbaum, 1987, 1988, 1991) can display the sensitivity of an inference to assumptions concerning an unobserved covariate. Thus, uncertainty about hidden bias can be described by considering what situations are consistent with the observed data but might lead to different inferences.

Recall the discussion of the unadjusted and adjusted effects. With no adjustment, Head Start participation appears to be detrimental in some ways; with adjustment, Head Start participation appears to have few discernible long-term effects. However, it is quite likely that hidden bias is present, i.e., that the Head Start group is disadvantaged in additional ways that were not measured. It is possible that the correct inference is that Head Start is beneficial with respect to some outcomes and that this could be demonstrated if it were possible to compare Head Start using similar study groups. Table F.8 presents a sensitivity analysis that considers this possibility.

The analysis displayed in Table F.8 assumes that after matching, there is hidden bias due to an imbalance in an unobserved covariate. The table gives the odds of matched subjects being in Head Start (or Model Head Start) due to an imbalance in the unobserved covariate that is consistent with the inference that Head Start (or Model Head Start) is significantly beneficial with respect to the given outcome. For example, consider *high school graduation* for the Florida Head Start versus No–Head Start comparison. If Head Start is beneficial with respect to this variable and if the hidden bias can be attributed to an imbalance in an unobserved covariate, the matched subjects must differ in their odds of being in Head Start by at least 50% owing to imbalances in the unobserved covariate. Similarly, if Head Start is beneficial with respect to *high school graduation* in Colorado, but hidden bias masks this result, the unobserved covariate could explain this situation if it tripled the odds of being in Head Start.

The results indicate that for the Florida Head Start versus No–Head Start comparison, the results are sensitive to small biases, i.e., it is plausible that Head Start is actually beneficial but that a relatively small imbalance in the study groups due to an unmeasured covariate could have led to the

Table F.8: Hidden Bias to Explain Benefits of Head Start

Variable	Florida Head Start vs. No–Head Start	Colorado Head Start vs. No–Head Start	Model vs. Regular Head Start
High school graduation	1.5[a]	3.0	3.0
Some postsecondary education	2.0	4.0	4.0
Ever on welfare	2.0	1.5	2.0
Employed or in school at interview	1.25	3.0	1.5
Teen parenthood	2.0	3.0	1.5
Education in years	2.0	5.0	3.0
GPA in elementary school	*	3.0	***
GPA in middle school	1.25	3.0	**
GPA in high school	1.25	4.0	***
Adult APL Survey score	2.0	4.0	4.0
Number of arrests	2.0	3.0	*
Number of convictions	1.5	4.0	*

Note. Values close to 1 show high sensitivity of inference to hidden bias; values 3 or above show insensitivity of inference to hidden bias.

[a] Odds of being in Head Start due to imbalance in unobserved covariate sufficient to infer that Head Start (or Model Head Start) is beneficial

* Significant benefit ($p < .05$) after matching adjustment.

** Significant benefit ($p < .01$) after matching adjustment.

*** Significant benefit ($p < .001$) after matching adjustment.

observed data. Still, for the Colorado Head Start versus No–Head Start comparison, the conclusion that there are no significant benefits of Head Start is relatively insensitive to moderately large hidden bias. This means that it is possible that Head Start is beneficial for the Colorado subjects, but the imbalance between groups not accounted for by the observed covariates would have to be fairly large. This is, however, plausible, since the Colorado study groups were quite different at base line.

For the Florida Model Head Start versus Regular Head Start comparison, the sensitivity to hidden bias varies according to the specific outcome. The GPA increases due to Model Head Start were highly significant, despite the hint of the Head Start parents' lower educational status at base line. The outcomes of more *employed or in school at time of interview* and less *teen parenthood* could be benefits of High/Scope if there is slight imbalance owing to an unmeasured covariate. However, a moderately large imbalance is necessary to explain the difference in outcomes for *high school graduation, some postsecondary education, years of education,* and *Adult APL Survey score.*

Another sensitivity analysis was conducted to assess uncertainty due to hidden bias resulting from the impact of an unmeasured covariate on the outcomes and on the treatment effect (Marcus, 1997). The results of this analysis for selected educational outcomes are presented on the next page, in

Table F.9: Treatment Effect Adjusted for Hidden Bias Under Selected Assumptions About an Unobserved Covariate

Variable	Florida Head Start vs. No–Head Start	Colorado Head Start vs. No–Head Start	Model vs. Regular Head Start
Education in years			
$\gamma_1 = .2\sigma, \gamma_2 = .2\sigma$ ($.2\sigma = .3$ year)	.22	-.02	.20
$\gamma_1 = .5\sigma, \gamma_2 = .5\sigma$ ($.5\sigma = .75$ year)	.58**	.28	.50
GPA in elementary school			
$\gamma_1 = .2\sigma, \gamma_2 = .2\sigma$ ($.2\sigma = .10$ point)	.28	.06	.90***
$\gamma_1 = .5\sigma, \gamma_2 = .5\sigma$ ($.5\sigma = .25$ point)	.51***	.29**	1.13***
GPA in middle school			
$\gamma_1 = .2\sigma, \gamma_2 = .2\sigma$ ($.2\sigma = .14$ point)	.31*	.00	.70***
$\gamma_1 = .5\sigma, \gamma_2 = .5\sigma$ ($.5\sigma = .35$ point)	.61***	.29+	1.00***
GPA in high school			
$\gamma_1 = .2\sigma, \gamma_2 = .2\sigma$ ($.2\sigma = .17$ point)	.40**	-.12	.88***
$\gamma_1 = .5\sigma, \gamma_2 = .5\sigma$ ($.5\sigma = .43$ point)	.79***	.27	1.31***

Note. γ_1 = the contribution to the outcome due to the difference between the most and least extreme effect of *u*. γ_2 = the contribution to the treatment effect due to the difference between the most and least extreme effect of *u*.

Table F.9. For example, if an extra ¾ year of education (about .5σ, a moderate effect) could be attributed to an unmeasured covariate *u*, and furthermore, if there is a treatment-group-by-*u* interaction effect such that the contribution to the treatment effect due to the difference between the most and least extreme level of the unobserved covariate is an additional ¾ year of education, then it is plausible that Head Start in Florida increases *years of education* by about ½ year. Results of this analysis also indicate that the existence of a moderately small hidden bias is consistent with a beneficial effect of Head Start in Florida with respect to *GPA*. It is plausible that a moderately large hidden bias is consistent with a significant increase in *years of education* for Head Start.

For Colorado, if the impact of an unmeasured covariate is moderately large, it seems possible that Head Start could raise *GPA* by about ¼ point. For the Model versus Regular Head Start comparison, hidden bias could mean that Model Head Start increases *years of education* by as much as ½ year and increases *GPA* by a full point.

This type of analysis indicates why tight experimental designs may be so crucial to identifying long-term program effects.

STUDY GROUP INTERACTIONS

Another planned question of interest for evaluating the long-term benefits of Head Start was the following: Does Head Start (or Model Head Start) have significant benefits for certain subpopulations? A look at those outcomes with significant propensity score quintile-by-study-group interactions can indicate whether Head Start (or Model Head Start) works better for subjects of differing socioeconomic levels. In addition, significant gender/race-by-study-group interactions can point to treatment effect differences among non–White females, White females, non–White males, White males. (See the footnote of Table F.1, p. 181, for an explanation of "White" and "non–White" in this study.)

To consider the question of whether Head Start (or Model Head Start) has different benefits for different socioeconomic levels, outcomes with significant propensity score quintile-by-study-group interactions were examined. (A look again at Table F.5 on p. 187 is helpful in thinking about this; recall that Q1 through Q5 give increasing levels of low socioeconomic status.) The highlights of this analysis (see Tables F.10 and F.11, pp. 196–197) are as follows: There is some evidence that Head Start in Colorado tended to be more beneficial for those subjects coming from the most severely disadvantaged homes. Those subjects in the Colorado Head Start who were of the lowest socioeconomic status were more likely to benefit from Head Start with respect to increasing *high school graduation* and avoiding being *ever on welfare*. Furthermore, there is evidence that in Colorado the effect of Head Start on increasing *GPA in elementary school* and increasing *years of education* was greater for subjects of relatively lower socioeconomic status. However, since the sample sizes within each subgroup are small, this analysis merely points to hypotheses that could be explored in future studies.

Analysis of the effects of Head Start with respect to race/gender differences revealed several possibilities for further research. In Florida, it appeared that Head Start was more beneficial for the African American females with respect to increasing *GPA in elementary school* and *middle school* ($p < .01$, $p < .05$, respectively). Generally, the White/other Head Start population in Florida tended to have lower GPAs and were more likely to be on welfare when compared with the White/other controls ($p < .05$, $p < .01$). This could be, at least in part, a reflection of a difference in Head Start recruiting for White/other versus African American children; however, this needs to be explored further.

OTHER METHODOLOGICAL CONSIDERATIONS

In the analysis of the LTBHS study, 12 outcomes were considered for each of three comparisons—Florida Head Start versus No–Head Start, Colorado Head Start versus No–Head Start, and Florida Model Head Start versus Regular Head Start. Multiple comparisons present the problem of the increased likelihood of incorrectly inferring that effects are statistically significant. Adjustments can be made for multiple comparisons (e.g., Bonferroni), but this approach would make it almost impossible to infer significance, and this seems too severe in this case.

Table F.10: Odds Ratios by Propensity Score Quintile

Variable/Quartile	Florida Head Start vs. No-Head Start	Colorado Head Start vs. No-Head Start	Model vs. Regular Head Start
High school graduation			
Q1	—	—	—
Q2	1.32	.50	2.60
Q3	.69	.21	.37
Q4	.70	7.71	.55
Q5	1.11	3.69	—
Some postsecondary education			
Q1	.84	.24	—
Q2	.70	.69	1.57
Q3	.70	.53	.20
Q4	1.10	—	.86
Q5	.83	.47	.20
Ever on welfare			
Q1	1.23	1.25	—
Q2	1.62	1.25	.87
Q3	.65	.47	.59
Q4	1.37	.18	.76
Q5	1.59	—	3.00
Employed or in school at time of interview			
Q1	2.16	.40	—
Q2	1.26	.43	2.93
Q3	3.11	2.22	.87
Q4	.44	.92	.83
Q5	2.60	—	.84
Teen parenthood			
Q1	—	1.75	—
Q2	1.53	1.72	.57
Q3	1.07	.83	.50
Q4	1.57	.13	.75
Q5	1.25	.55	—

One approach to this dilemma is to consider a context-specific interpretation (Marcus, 1993). Since the multiple comparisons were planned before carrying out the analysis, no adjustment was made for multiplicity in this analysis. However, suffice it to say that for a type I error $\alpha = .05$, it would not be surprising if several of the many comparisons made during the analysis reached statistical significance by chance. However, consistent inferences, e.g., the apparent benefit of Model (High/Scope) Head Start in elementary school, middle school, and high school, probably reflect genuine effects. Also, these hypotheses were derived from existing findings.

Another consideration is the nature of interview data. The estimates of the benefits of Head Start may be subject to recall bias if the accuracy of the interview data differs between the study groups. Unfortunately, this is difficult to determine, but there is no reason to believe that the recall of the

Table F.11: Treatment Effects by Propensity Score Quintile

Variable/Quartile	Florida Head Start vs. No–Head Start	Colorado Head Start vs. No–Head Start	Model vs. Regular Head Start
Education in years			
Q1	.03	-1.12	-.07
Q2	-.11	-.37	.60
Q3	-.46	-.57	-.99
Q4	-.41	.82	.07
Q5	-.13	.26	-1.27
GPA in elementary school			
Q1	-.05	-.20	.52
Q2	.30	-.01	.83
Q3	-.13	-.24	.72
Q4	.26	-.13	.77
Q5	.12	.45	.68
GPA in middle school			
Q1	-.02	-.55	-.10
Q2	.22	-.54	1.02
Q3	-.12	.18	.30
Q4	.18	.16	.62
Q5	-.07	-1.56	.08
GPA in high school			
Q1	.09	-.42	.75
Q2	.11	-.78	1.14
Q3	.16	-.26	.18
Q4	-.02	.42	.66
Q5	.02	-1.02	.35
Adult APL Survey score			
Q1	-2.63	-1.28	-13.15
Q2	-.10	-5.12	2.10
Q3	.42	.17	-1.33
Q4	-4.38	-1.23	.03
Q5	-3.19	8.24	-8.81
Number of arrests			
Q1	.31	-.96	1.40
Q2	-.15	4.91	-2.48
Q3	1.24	1.34	.60
Q4	-.94	-3.79	-.34
Q5	.37	.98	-2.00
Number of convictions			
Q1	.27	-.08	-1.17
Q2	-.17	2.34	-2.77
Q3	.90	1.02	.13
Q4	-1.03	-1.49	-.08
Q5	.44	1.14	-1.5

BEST COPY AVAILABLE

treated subjects concerning their status as children differed from that of the control subjects.

Missing ("unknown" response) data can also lead to bias in estimating treatment effects if the reason for the missing data is related to the study group and the outcome. To explore this possibility informally, I used propensity scores to compare subjects who had missing data for a particular outcome with subjects who had data for the given outcome. This analysis revealed a possible bias due to missing data for the Colorado *GPA in elementary school* and *middle school*. Subjects with missing data appeared to be of relatively higher socioeconomic status ($p < .05$ for elementary school and $p < .05$ for middle school). Since a somewhat smaller proportion of the Head Start group than of the control group had missing data on these two variables (17% vs. 22%, 25% vs. 29%), this could lead to a slight bias in favor of the control group. Thus Head Start might be slightly more beneficial than the existing data shows, and a revised inference would be that there is little difference between the Colorado groups with respect to *GPA*.

CONCLUSIONS

Assessing the long-term benefits of these Head Start programs is difficult, because it seems likely that the most severely disadvantaged children were recruited to the Head Start programs, leaving a pool of potential control candidates who were different at base line. Analysis of the study groups revealed that compared with the control groups at both sites, the Head Start groups *did* seem to be from families with relatively less education and more poverty. Since the status of the children before entering Head Start appears to be a stronger predictor of long-term outcomes than Head Start attendance is, it is essential that estimates of the long-term benefits of Head Start reflect baseline imbalances in socioeconomic status. Adjustments for observed covariates reduce selection bias in the estimates; however, a thorough analysis involves uncertainty regarding unobserved study group differences.

After adjusting for measured study group differences at base line, there appears to be little clear-cut evidence of long-term benefits of Head Start versus No–Head Start, for both Florida and Colorado. Yet, there is evidence of highly significant benefits of Model (High/Scope) Head Start over Regular Head Start with respect to increased *GPA in elementary school, middle school,* and *high school,* as well as with respect to lower *numbers of arrests* and *convictions.*

The sensitivity analyses show that for the Florida Head Start versus No–Head Start comparison, it is quite possible that a relatively small hidden bias due to an imbalance in an unmeasured covariate (e.g., parents' attitudes toward education) is masking the long-term benefits of Head Start. For the Colorado comparison, it is plausible that Head Start has long-term benefits if a moderately large imbalance with respect to an unobserved covariate is present; given the large imbalance in the Colorado study groups with respect to measured baseline covariates, this is indeed a possibility. The Florida Model Head Start versus Regular Head Start comparison shows that there is strong evidence that High/Scope Head Start helps to increase *GPA in elementary school, middle school,* and *high school.* There is also evidence of reduced *numbers of arrests* and *convictions* for the Model Head Start subjects. Furthermore, sensitivity analysis shows the possibility of reductions in *teen parenthood* and *unemployment* that are consistent with the existence of a small hidden bias.

In summary, there is evidence that Model Head Start produces long-term benefits of increased GPAs and reduced crime; however, the analysis of the Long-term Benefits of Head Start study cannot demonstrate with certainty that there are any long-term benefits of Head Start over No–Head Start. Nevertheless, sensitivity analyses show that it is possible that Head Start is significantly beneficial but that selection bias may account for the ostensible lack of benefits.

References

Campbell, D., & Boruch, R. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. Bennett and A. Lumsdaine (Eds.), *Evaluation and experiment* (pp. 195–296). New York: Academic.

Cochran, W. G. (1965). The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society,* Series A, *128,* 234–255.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24,* 205–213.

Cox, D. R. (1970). *Analysis of binary data.* London: Chapman and Hall.

Fisher, R. A. (1935). *The design of experiments.* London: Hafner.

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics, 2*(4), 405–420.

Marcus, S. M. (1993). Multiplicity issues in clinical trials: Context-specific interpretations. *Proceedings of the American Statistical Association, August 1993.*

Marcus, S. M. (1997). Using omitted variable bias to assess uncertainty in the estimation of an AIDS education treatment effect. *Journal of Educational and Behavioral Statistics, 22*(2), 193–201.

Oden, S. (1995, January). *The Long-Term Benefits of Head Start study: A background overview and current status* (technical report). Ypsilanti, MI: High/Scope Educational Research Foundation.

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika, 74*(1), 13–26.

Rosenbaum, P. R. (1988). Sensitivity analysis for matching with multiple controls. *Biometrika, 75*(3), 577–581.

Rosenbaum, P. R. (1989). Optimal matching of observational studies. *Journal of the American Statistical Association, 84,* 1024–1032.

Rosenbaum, P. R. (1991). Sensitivity analysis for matched case-control studies. *Biometrics, 47*(1), 87–100.

Rosenbaum, P. R. (1995). *Observational studies.* New York: Springer-Verlag.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rosenbaum, P., & Rubin, D. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79,* 516–524.

References

Abt Associates, Inc., with Yale University, The Bush Center in Child Development and Social Policy. (1995). *Review of Research on Supportive Interventions for Children and Families* (Vols. 1 and 2). Washington, DC: Administration on Children, Youth and Families, U.S. Department of Health and Human Services.

American College Testing Program. (1976). *User's guide: Adult APL Survey.* Iowa City, IA: Author.

American Council on Education (1995). *Thirteenth Annual Status Report on Minorities in Higher Education.* Washington, DC: Author.

Barnett, W. S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources, 27*(2), 279–312.

Barnett, W. S. (1995, Winter). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25–50.

Barnett, W. S. (1996). *Lives in the balance: Age-27 benefit-cost analysis of the High/Scope Perry Preschool program* (Monographs of the High/Scope Educational Research Foundation, 11). Ypsilanti, MI: High/Scope Press.

Bee, C. K. (1981). *A longitudinal study to determine if Head Start has lasting effects.* Ann Arbor, MI: UMI Dissertation Services.

Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19* (Monographs of the High/Scope Educational Research Foundation, 8). Ypsilanti, MI: High/Scope Press.

Bissell, J. S. (1971). *Implementation of planned variation in Head Start.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Child Development.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology, 99,* 353–395.

Brooks-Gunn, J., Phelps, E., & Elder, G. H., Jr. (1991). Studying lives through time: Secondary analyses in developmental psychology. *Developmental Psychology, 27,* 899–910.

Brown (1978). Long-term gains from early intervention: An overview of current research. In B. Brown (Ed.), *Found: Long-term gains from early intervention* (pp. 169–185). Boulder, CO: Westview Press.

Bryant, D. M., Lau, L. B., Burchinal, M., & Sparling, J. J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9,* 289–310.

Campbell, D. T., & Boruch, R. F. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett & A. Lumsdaine (Eds.), *Evaluation and experiments: Some critical issues in assessing social programs* (pp. 195–296). New York: Academic Press.

Campbell, D. T., & Erlebacher, A. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: A national debate* (Vol. 3, pp. 321–342). New York: Brunner/Mazel.

Campbell, D. T., & Frey, P. W. (1970). The implications of learning theory for the fade-out of gains from compensatory education. In J. Hellmuth (Ed.), *Compensatory education: A national debate* (Vol. 3, pp. 185–223). New York: Brunner/Mazel.

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*, 684–698.

Campbell, F. A., & Ramey, C. T. (1995, Winter). Cognitive and school outcomes for high-risk African American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32*, 743–772.

Campbell, R. T., & Parker, R. N. (1983). Substantive and statistical considerations in the interpretation of multiple measures of SES. *Social Forces, 62*, 450–466.

Cawley, J. F., & Goodstein, H. A. (1966). *An assessment of intelligence, psycholinguistic abilities, and learning aptitudes among preschool children.* Storrs, CT: Univerity of Connecticut (ERIC Document Reproduction Service No. ED 014 323).

Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., & Phillips, D. A. (1991). Children of the National Longitudinal Survey of Youth: A unique research opportunity. *Developmental Psychology, 27*, 918–931.

Clarke, S. H., & Campbell, F. A. (1997, April). *The Abecedarian Project and youth crime.* Paper presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.

Consortium for Longitudinal Studies. (1983). *As the twig is bent . . . Lasting effects of preschool programs.* Hillsdale, NJ: Lawrence Erlbaum.

Copple, C. E., Cline, M. G., & Smith, A. N. (1987). *Path to the future: Long-term effects of Head Start in the Philadelphia School District.* Washington, DC: U.S. Department of Health and Human Services.

CSR, Inc. (1985). *An annotated bibliography of the Head Start research since 1965. Head Start Evaluation, Synthesis, and Utilization Project.* Washington, DC: U.S. Government Printing Office.

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *The American Economic Review, 85*, 341–364.

Currie, J. M. (1997, Summer/Fall). Choosing among alternative programs for poor children. *The Future of Children, 7*(2), 113–131.

Darlington, R. B. (1991). The long-term effects of model preschool programs. In L. Okagaki, R. J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking* (pp. 203–215). Hillsdale, NJ: Lawrence Erlbaum.

Datta, L., McHale, C., & Mitchell, S. (1976). *The effects of Head Start classroom experience on some aspects of child development: A summary report of national evaluations, 1966–1969* (DHEW Publication No. OHD-76-30088). Washington, DC: U.S. Government Printing Office.

Duncan, G. J. (1991). Made in heaven: Secondary data analysis and interdisciplinary collaborators. *Developmental Psychology, 27*, 949–951.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development, 65*, 296–318.

Duncan, G. J., with Coe, R. D., Corcoran, M. E., Hill, M. S., & Hoffman, S. D. (1984). *Years of poverty, years of plenty. The changing economic fortunes of American workers and families.* Ann Arbor, MI: Institute for Social Research, University of Michigan.

Duncan, O. D. (1961). A socioeconomic index for all occupations. In Reiss, A. J., Jr. (Ed.), *Occupations and social status* (pp. 109–138). New York: The Free Press.

Dunst, C. J., & Trivett, C. M. (1994). Methodological considerations and strategies for studying the long-term effects of early intervention. In S. L. Friedman & H. C. Haywood (Eds.), *Developmental follow-up. Concepts, domains, and methods* (pp. 277–313). San Diego: Academic Press.

Elder, G. H., Jr. (1974). *Children of the Great Depression.* Chicago: University of Chicago Press.

Ellsworth Associates, Inc. (1995). *Head Start research from 1985–1995: An annotated bibliography.* Washington, DC: Administration on Children, Youth and Families, U.S. Department of Health and Human Services.

Emrick, J. A., Sorenson, P. H., & Stearns, M. S. (1973). *Interim evaluation of the National Follow Through Program 1969–1971.* Menlo Park, CA: SRI International.

Epstein, A. S., Schweinhart, L. J., & McAdoo, L. (1996). *Models of early childhood education.* Ypsilanti, MI: High/Scope Press.

Fosburg, L., & Brown, B. (1984). *The effects of Head Start health services: Executive summary of the Head Start evaluation.* Cambridge, MA: Abt Associates.

Frede, E. C. (1995, Winter). The role of program quality in producing early childhood program benefits. *The Future of Children, 5*(3), 115–132.

Friedman, S. L., Haywood, H. C., & Livesey, K. (1994). From the past to the future of developmental follow-up research. In S. L. Friedman & H. C. Haywood (Eds.), *Developmental follow-up. Concepts, domains, and methods* (pp. 3–26). San Diego: Academic Press.

Fuerst, J. S., & Fuerst, D. (1993, April). Chicago experience with an early childhood program: The special case of the Child-Parent Center program. *Urban Education, 28*(1), 69–96.

Gamble, T., & Zigler, E. (1989). The Head Start Synthesis Project: A critique. *Journal of Applied Developmental Psychology, 10*, 267–274.

Garber, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk.* Washington, DC: American Association on Mental Retardation.

Gersten, R., & Keating, T. (1987). Improving high school performance of "at-risk" students: A study of long-term benefits of direct instruction. *Educational Leadership, 44*(6), 28–31.

Gomby, D. S., & Larner, M. B. (Eds.). (1995). *The Future of Children, 5*(3).

Gomby, D. S., Larner, M. B., Stevenson, C. S., Lewit, E. M., & Behrman, R. E. (1995, Winter). Long-term outcomes of early childhood programs: Analysis and recommendations. *The Future of Children, 5*(3), 6–24.

Gotts, E. E. (1989). *HOPE, preschool to graduation: Contributions to parenting and school-family relations theory and practice.* Charleston, WV: Appalachia Educational Laboratory.

Gray, S. W., Ramsey, B. K., & Klaus, R. A. (1982). *From 3 to 20: The Early Training Project.* Baltimore, MD: University Park Press.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361–396.

Hale, B. A., Seitz, V., & Zigler, E. (1990). Health services and Head Start: A forgotten formula. *Journal of Applied Developmental Psychology, 11,* 447–458.

Hebbeler, K. M. (1985). An old and a new question on the effects of early education for children from low-income families. *Educational Evaluation and Policy Analysis, 7*(3), 207–216.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Hill, M. S. (1992). *The Panel Study of Income Dynamics. A user's guide.* Newbury Park, CA: Sage, Inc.

Hohmann, M., Banet, B., & Weikart, D. P. (1979). *Young children in action: A manual for preschool educators.* Ypsilanti, MI: High/Scope Press.

Hohmann, M., & Weikart, D. P. (1995). *Educating young children: Active learning practices for preschool and child care programs.* Ypsilanti, MI: High/Scope Press.

Irvine, D. J. (1982). *Evaluation of the New York State Experimental Prekindergarten Program.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Johnson, D. L. (1988). Primary prevention of behavior problems in young children: The Houston Parent-Child Development Center. In R. H. Price, E. L. Cowen, R. P. Lorion, & J. Ramos-McKay (Eds.), *Fourteen ounces of prevention: A casebook for practitioners* (pp. 44–52). Washington, DC: American Psychological Association.

Kassebaum, N. L. (1994). Head Start: Only the best for America's children. *American Psychologist, 49*(2), 123–126.

Kennedy, M. M. (1978, June). Findings from the Follow Through Planned Variation study. *Educational Researcher, 7*(6), 3–11.

Lally, J. R., Mangione, P. L., & Honig, A. S. (1988). The Syracuse University Family Development Research Program: Long-range impact of an early intervention with low-income children and their families. In D. R. Powell (Ed.), *Annual Advances in Applied Developmental Psychology: Vol. 3. Parent education as early childhood intervention: Emerging directions in theory, research, and practice* (pp. 79–104). Norwood, NJ: Ablex.

Lansing, J. B., & Morgan, J. N. (1971). *Economic survey methods.* Ann Arbor, MI: Institute for Social Research, University of Michigan.

Lazar, I., & Darlington, R. B. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47*(2–3, Serial No. 195).

Lee, V. E., Brooks-Gunn, J., & Schnur, E. (1988). Does Head Start work? A one-year follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Developmental Psychology, 24,* 210–222.

Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F.-R. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61,* 495–507.

Lee, V. E., & Loeb, S. (1995, Spring). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis, 17*(1), 62–82.

Levenstein, P., O'Hara, J., & Madden, J. (1983). The Mother-Child Home Program of the Verbal Interaction Project. In Consortium for Longitudinal Studies, *As the twig is bent . . . Lasting effects of preschool programs* (pp. 237–263). Hillsdale, NJ: Lawrence Erlbaum.

Lewit, E. M., Terman, D. L., & Behrman, R. E. (1997, Summer/Fall). Children and poverty: Analysis and recommendations. *The Future of Children, 7*(2), 4–24.

Love, J. M., & Grover, J. (1987). *Study of Head Start recruitment and enrollment: Final report.* Washington, DC: Administraton on Children, Youth and Families.

Lukas, C. V., & Wohlleb, C. (1972). *Implementation of Head Start Planned Variation: 1970–71* (Parts I & II). Cambridge, MA: Huron Institute.

Mann, A. J., Harrell, A. V., & Hurt, M. (1978). A review of Head Start research since 1969. In B. Brown (Ed.), *Found: Long-term gains from early intervention* (pp. 129–168). Boulder, CO: Westview Press.

Manski, C. F. (1995). *Identification problems in the social sciences.* Cambridge, MA: Harvard University Press.

Manski, C. F., & Garfinkel, I. (1992). Introducation. In F. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs* (pp. 1–22). Cambridge, MA: Harvard University Press.

McCall, R. B., & Appelbaum, M. I. (1991). Some issues of conducting secondary analyses. *Developmental Psychology, 27*, 911–917.

McKey, R. H., Condelli, L., Ganson, H., Barrett, B. J., McConkey, C., & Plantz, M. C. (1985, June). *The impact of Head Start on children, families, and communities: Head Start Synthesis Project* (Final Report of the Head Start Evaluation, Synthesis, and Utilization Project, CSR, Inc.) Washington, DC: U.S. Government Printing Office.

Miller, L. B., & Bizzell, R. P. (1984). Long-term effects of four preschool programs: Ninth- and tenth-grade results. *Child Development, 55*, 1570–1587.

Monroe, E., & McDonald, M. S. (1981). *A follow-up study of the 1966 Head Start program. Rome City Schools, Rome, Georgia.* Unpublished paper.

Oden, S., Kelly, M. A., Ma, Z., & Weikart, D. P. (1992). *Challenging the potential: Programs for talented disadvantaged youth.* Ypsilanti, MI: High/Scope Press.

Osborn, A. F. (1987). Assessing the socioeconomic status of families. *Sociology, 21*, 429–448.

Palmer, F. H. (1983). The Harlem study: Effects by type of training, age of training, and social class. In Consortium for Longitudinal Studies, *As the twig is bent . . . Lasting effects of preschool programs* (pp. 201–236). Hillsdale, NJ: Lawrence Erlbaum.

Phillips, D. A., & Cabrera, N. J. (Eds.). (1996). *Beyond the blueprint: Directions for research on Head Start's families.* Washington, DC: National Academy Press.

Powell, G., & Roane, M. (1996, Winter). The fade-out effect is a fading argument. *National Head Start Journal*, pp. 25–33.

Ramey, C. T., Bryant, D. M., Campbell, F. A., Sparling, J. J., & Wasik, B. H. (1988). In R. H. Price, E. L., Cowen, R. P. Lorian, & J. Ramos-McKay (Eds.), *Fourteen ounces of prevention: A casebook for practitioners* (pp. 32–43). Washington, DC: American Psychological Association.

Ramey, C. T., Bryant, D. M., & Suarez, T. M. (1985). Preschool compensatory education and modifiability of intelligence: A critical review. In D. Detterman (Ed.), *Current topics in intelligence* (pp. 247–296). Norwood, NJ: Ablex

Reynolds, A. J. (1992). Mediated effects of preschool intervention. *Early Education and Development, 3*, 139–164.

Reynolds, A. J., Chang, H., & Temple, J. A. (1997, April). *Early educational intervention and juvenile delinquency: Findings from the Chicago longitudinal study.* Paper presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.

Rog, D. J. (1994). Expanding the boundaries of evaluation: Strategies for refining and evaluating ill-defined interventions. In S. L. Friedman & H. C. Haywood (Eds.), *Developmental follow-up. Concepts, domains, and methods* (pp. 139–154). San Diego: Academic Press.

Ross, R. T., Begab, M. J., Dondis, E. H., Giampiccolo, J. S., & Meyers, C. E. (1985). *Lives of the mentally retarded. A forty-year follow-up study.* Stanford, CA: Stanford University Press.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P., with Barnett, W. S., & Epstein, A. S. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27* (Monographs of the High/Scope Educational Research Foundation, 10). Ypsilanti, MI: High/Scope Press.

Schweinhart, L. J., & Wallgren, C. R. (1993). Effects of a Follow Through program on school achievement. *Journal of Research in Childhood Education, 8,* 43–56.

Schweinhart, L. J., & Weikart, D. P. (1997). Lasting differences: *The High/Scope Perry Preschool Curriculum Comparison study through age 23* (Monographs of the High/Scope Educational Research Foundation, 12). Ypsilanti, MI: High/Scope Press.

Schweinhart, L. J., Weikart, D. P., & Larner, M. B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly, 1,* 15–45.

Seitz, V., Apfel, N. H., & Efron, C. (1978). Long-term effects of early intervention: The New Haven Project. In B. Brown (Ed.), *Found: Long-term gains from early intervention* (pp. 79–110). Boulder, CO: Westview Press.

Seitz, V., Apfel, N. H., Rosenbaum, L. K., & Zigler, E. (1983). Long-term effects of Projects Head Start and Follow Through: The New Haven Project. In Consortium for Longitudinal Studies, *As the twig is bent . . . Lasting effects of preschool programs* (pp. 299–332). Hillsdale, NJ: Lawrence Erlbaum.

Seitz, V., Rosenbaum, L. K., & Apfel, N. H., (1985). Effects of family support intervention: A ten-year follow-up. *Child Development, 56,* 376–391.

Smith, M. S. (1973). *Some short-term effects of Project Head Start: A preliminary report on the second year of planned variation, 1970–71.* Cambridge, MA: Huron Institute.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model* (Volume IV-A, An evaluation of Follow Through). Washington, DC: U.S. Office of Education.

Stevens, G., & Featherman, D. L. (1981). A revised socioeconomic index of occupational status. *Social Science Research, 10,* 364–395.

Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale. Manual for the Third Revision. Form L–M.* Boston: Houghton Mifflin.

U. S. Bureau of the Census. (1971). *1970 Census of Population, Alphabetical index of industries and occupations.* Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1972). *Census of population and housing: 1970 Census tracts, Ann Arbor, Mich., Standard Metropolitan Statistical Area.* [Final report PH(1)-11]. Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1989). *Studies in marriage and the family. Singleness in America. Married-couple families with children.* (Current Population Reports, Series P-23, No. 162). Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1994). *School enrollment—Social and economic characteristics of students* (Current Population Reports, Series P-20, No. 479). Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services (USDHHS). (1990, September). *Head Start research and evaluation: A blueprint for the future.* Washington, DC: Author.

U.S. Department of Health and Human Services (USDHHS). (1993). *Creating a 21st century Head Start: Final report of the Advisory Committee on Head Start Quality and Expansion.* Washington, DC: Author.

U.S. Department of Health and Human Services (USDHHS). (1998a) *Head Start Act* (October 27). Washington DC: Author.

U.S. Department of Health and Human Services (USDHHS). (1998b) *Head Start 1998 Statistical Fact Sheet.* Washington DC: Head Start Bureau, Administration on Children, Youth and Families.

U.S. Federal Register: April 22, 1996. (Vol. 61, No. 78), Part II. Head Start Program; Proposed Rules. Pages 17753–17792. Washington, DC: Department of Health and Human Services, Administration for Children and Families.

U.S. Federal Register: November 5, 1996. (Vol. 61, No. 215), Part III. Head Start Program; Final Rules. Pages 57186–60106. Washington, DC: Department of Health and Education, Administration for Children and Families.

U.S. General Accounting Office. (1997, April). *Head Start: Research provides little information on impact of current programs* (Report No. GAO/HEHS-97-59). Washington, DC: Author.

Walberg, H. J. (1992). Potential benefits of longitudinal studies in education. In E. A. Ramp (Ed.), *Follow-Through program and policy issues* (pp. 87–105). Washington, DC: U.S. Department of Education, Office of Research, Office of Educational Research and Improvement .

Washington, V., & Oyemade, U. S. (1987). *Project Head Start: Past, present, & future trends in the context of family needs.* New York: Garland.

Weisberg, H. I. (1973). *Short-term cognitive effects of Head Start programs: A report on the third year of planned variation—1971–72.* Cambridge, MA: Huron Institute. (ERIC Document Reproduction Service No. ED 093 047)

Werner, E. E., & Smith, R. S. (1977). *Kauai's children come of age.* Honolulu: University Press of Hawaii.

Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible.* New York: McGraw-Hill.

Westinghouse Learning Corporation. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (Vols. 1–2, Westinghouse Learning Corporation and Ohio University report). Washington, DC: Clearinghouse for Federal, Scientific, & Technical Information.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy.* Chicago: University of Chicago Press.

Zigler, E. (1993, July 24). Head Start, the whole story. *The New York Times,* OP-ED page.

Zigler, E., & Muenchow, S. (1992). *Head Start: The inside story of America's most successful educational experiment.* New York: Basic Books.

Zigler, E. & Styfco, S. J. (1994). Head Start: Criticisms in a constuctive context. *American Psychologist, 49*(2), 127–132.

Zigler, E., & Valentine, J. (Eds.). (1979). *Project Head Start: A legacy of the War on Poverty.* New York: Free Press.

Commentaries

A Head Start—For Life?

Suzanne M. Randolph
Associate Professor
Department of Family Studies
University of Maryland—College Park

STUDYING THE IMPACT OF HEAD START THEN AND NOW

Since Head Start was created in 1965, researchers have attempted to document its long-term impact on the children and families who participate in the program. Although early research attempts, specifically the Westinghouse Learning Corporation study (1969), would not withstand a test of scientific rigor in today's evaluation climate (Rossi, Freeman, & Lipsey, 1999; U.S. General Accounting Office [GAO], 1997), the data from this research were often used either to support the immediate benefits of Head Start or to undermine the potential long-lasting benefits of Head Start. Also, it should be noted that "long-lasting," as used in the evaluation of Head Start and Head Start–like programs, meant extending only into the elementary school years, with a few exceptions (e.g., Berrueta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984; Schweinhart, Barnes, & Weikart, 1993; for reviews, see Barnett, 1995; Yoshikawa, 1995). Over time, the precision and specificity with which the research community is able to measure the impact of Head Start has improved. We now have more reliable and valid measurement strategies, enhanced study designs, and more sophisticated data analysis techniques. The Long-Term Benefits of Head Start (LTBHS) study has taken advantage of all these advances. It is an investigation that provides partial empirical support for the long-term impact of Head Start as well as guidance for the next generation of evaluation and research on Head Start's impact.

In the 1990s, three major advisory groups convened in an effort to provide guidance to the federal government and to the field in conducting research on Head Start children and families: the panel that produced *A Blueprint for the Future* (U.S. Department of Health and Human Services [USDHHS], 1990), the panel that produced the *Creating a 21st Century Head Start* report (USDHHS, 1993), and the National Academy of Sciences Head Start Research Roundtable, which produced the *Beyond the Blueprint* report (Phillips & Cabrera, 1996). In the documents summarizing the work of each of these advisory groups, the chief recommendation was the expansion of the research agenda in order to capitalize on Head Start's value as a "national laboratory." During this period, the Head Start Bureau also convened a technical working group to provide guidance on updating the Head Start program performance measures, which at the time were more than 20 years old. The resulting documents, *Charting Our Progress* (USDHHS, 1995) and the *Head Start Program Performance Standards* (U.S. Federal Register, 1996) and other regulations, provide guidance on developing, implementing, and monitoring quality early childhood education programs in Head Start.

The growing evidence of the role of quality in producing early childhood program benefits (Frede, 1995; Love, 1997; National Institute of Child Health and Development [NICHD] Early Child Care Research Network, 1999) has further fueled researchers' interest in the short- and long-term impact that programs like Head Start have on young children's development. The Head Start Quality Research Centers established by the Administration for Children, Youth and Families and the Family and Child Experiences Survey (McKey, Tarullo, & Doan, 1999; Zill, Resnick, & McKey, 1999) funded by the Head Start Bureau are further indications of the federal government's commitment to better understand the impact of Head Start on children and families. Additional impetus for this research was provided by the *National Goals 2000 Panel Report* (U.S. Department of Education, 1995).

THE CONTRIBUTIONS OF THE LTBHS STUDY

The Long-Term Benefits of Head Start study is based on an initial design and sampling plan determined long before any of these initiatives of the 1990s, but through the creativity and inventiveness of the investigative team, it developed into a study that can offer guidance to researchers planning future studies. Despite its limitations with respect to generalizability, the LTBHS study has the potential to inform a twenty-first century Head Start research agenda in terms of addressing several design, measurement, and data analysis issues.

Design Issues

Program variation is key in Head Start. Programs vary, for example, in terms of their geographic location (rural, urban), their target population's racial/ethnic composition, and their educational model. The LTBHS study faced its own challenges in trying to respond to such differences, even though there was planned variation. Notably, to understand the long-term impact of programs directed at target populations that differed widely both geographically and racially/ethnically, the LTBHS study did not use a race-comparative or geographic-comparative design. It has long been argued that race-comparative paradigms involving under-studied populations yield little in the way of advancing the knowledge base. Researchers planning future studies of program impact should follow the LTBHS study's approach in evaluating the long-term impact for specific racial/ethnic or geographically different groups. The growing numbers of ethnic minorities, particularly those of immigrant status, is likely to result in a service population that will require that we adopt such approaches to our work.

Also, program quality will continue to matter, and as we educate the public about what matters in early childhood education, parents and policymakers will demand that our research help them understand the relationship between program quality and later child outcomes. Toward that end, we will need more studies like the LTBHS study—ones that are longitudinal in nature, extend beyond the elementary school years, and incorporate sufficient information about the nature of the programs studied. We need to know particularly about the quality-related aspects of programs so that readers can interpret child outcomes in light of programmatic and implementation issues.

The LTBHS study highlights both the promise and the challenge of conducting longitudinal Head Start research. The promise is that through well-planned and conceptualized studies, it is possible to document long-term benefits of Head Start through its participants' young-adult years. The challenge, however, lies in the need to plan for follow-up—to devise incentives to minimize attrition and to establish relationships with families and agencies that will endure well beyond the preschoolers' years in Head Start.

Measurement Issues

The measurement of life events over time in the LTBHS study—in terms of both children's later adversities and their resiliency—is noteworthy. While some measures are proxies for other constructs, the variables used in this study provide guidance in terms of what proximal or distal factors may provide indicators of long-lasting benefits of Head Start. Research such as that recommended recently by the National Academy of Sciences panel on data and methods to assess the impact of welfare reform (Moffitt & Ver Ploeg, 1999) could benefit from considering using the LTBHS study's variables in conducting studies that involve both children and their parents.

The LTBHS study's use of existing automated, linked public administrative data systems also provides lessons for the field. The study points out the limitations as well as the value of using these systems to collect retrospective data. In the future, ethnographic or other qualitative data should be considered by researchers to provide a more complete context for understanding both the significant and the nonsignificant findings. Such data could also be helpful in identifying or ruling out rival explanations for the results.

Data Analysis Issues

It is laudatory and probably unusual that the LTBHS study employed two independent statisticians to apply different analytical strategies to the data. Although some of Xie's and Marcus's findings are parallel, others are unique to the technique applied. There has been an increase in the use of public-use data sets (such as the data set from the National Longitudinal Study of Youth) to answer complex questions about the lives of children over time, especially across childhood, adolescence, and young adulthood. Related to Head Start, on the horizon is the Early Childhood Longitudinal Study–Birth Cohort (ECLS–B) sponsored by the National Center for Education Statistics. It will provide data from birth to age 3 to complement other public-use data sets focused on care and education of children of kindergarten age and older. Data from the Early Head Start Research and Evaluation Project are also forthcoming. The LTBHS study points out what is possible and why we need multiple approaches to a research question, even when the same data set is being used. As data sets from the ECLS–B, the Early Head Start Research and Evaluation Project, and the planned national evaluation of Head Start become available for use by the public, researchers planning their analytic strategies should consider the data analysis lessons learned from the LTBHS study.

WHERE DO WE GO FROM HERE?

In the 1998 Head Start Act, Congress included an amendment requiring the Secretary of the Department of Health and Human Services to convene an advisory committee to make recommendations on the design of a national Head Start impact research study. This committee is charged with recommending to the secretary the design, measures, sampling plan, and other aspects of the research necessary to generate a national estimate of the difference that Head Start makes in a child's life. This legislation was passed in a climate that requires federal agencies to document results of their funding as legislated by the 1993 Government Performance and Results Act (GPRA)(U.S. GAO, 1997, 1998; U.S. Office of Management and Budget, 1994). The LTBHS study provides estimates of some program effects in local sites; however, these findings are far from being an estimate of Head Start's national impact. Nonetheless, the lessons learned from the LTBHS study about collecting data longitudinally—for diverse samples in different regions and for a variety of outcome measures—are noteworthy.

Despite the contributions of the LTBHS study, several political and empirical constraints exist in the current climate. Policy changes such as welfare reform limit poor mothers' use of part-day Head Start programs and other Head Start–like programs, which are likely to be child care settings of higher quality than most. Also, the implementation of universal prekindergarten in some states will make it difficult to meet the real-world challenges of finding a comparison group suitable for research investigating whether Head Start makes a difference. Research will be further complicated by the joint initiative of the Head Start and Child Care bureaus that provide Head Start–like services to children in other child care settings and/or enhance quality (e.g., through provider training) in non–Head Start settings.

A twenty-first century Head Start agenda will require that we find solutions to the dilemma of needing scientific rigor while implementing studies within real-world constraints. The LTBHS study offers some solutions, but it is only a starting point. Head Start program variation would continue to pose a challenge even if all the resources in the world were available. Even more so, we need theory-based research that goes beyond Bronfenbrenner's classic conceptualization of children's and families' ecologies. One promising model is the integrative model of development of children of color (Garcia-Coll, Lamberty, Jenkins, Crnic, & McAdoo, 1996). This framework builds on Bronfenbrenner's theory by including several concepts often ignored in research on young children—concepts such as social position, racism, discrimination, the promoting and inhibiting nature of child care and early childhood education environments, and the adaptive cultures of racial/ethnic minority groups that have developed in response to these factors. Yet, we recognize that the impact of early childhood education is an issue not only for low-income families or families of color but also for middle- and upper-class parents and for White parents. Therefore, as the public investment in early childhood programs increases, the challenge will remain to translate research findings into large-scale public programs and policy.

Who knows, perhaps our advances in evaluation generally and in early childhood education research more specifically will take us into a twenty-first century evaluation and research agenda for Head Start and Head Start–like programs that will indeed enable us to answer this question: Will we give all children a Head Start—for life?

References

Barnett, W. S. (1995, Winter). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25–50.

Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19* (Monographs of the High/Scope Educational Research Foundation, 8). Ypsilanti, MI: High/Scope Press.

Frede, E. C. (1995, Winter). The role of program quality in producing early childhood program benefits. *The Future of Children, 5*(3), 115–132.

Garcia-Coll, C., Lamberty, G., Jenkins, R., Crnic, D., & McAdoo, H. (1996). An integrative model of developmental competencies in children of color. *Child Development, 67*, 1891–1914.

Love, J. M. (1997). Quality in child care centers. *Early Childhood Research & Policy Briefs, 1*(1), 1–4.

McKey, R. H., Tarullo, L. B., & Doan, H. M. (1999). *FACES: The Head Start Family and Child Experiences Survey.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Albuquerque, NM, April.

Moffitt, R. A., & Ver Ploeg, M. (Eds.). (1999). *Evaluating welfare reform: A framework and review of current work* (Interim report). Washington, DC: National Academy Press.

National Institute of Child Health and Development (NICHD) Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health, 89*(7), 1072–1077.

Phillips, D. A., & Cabrera, N. J. (1996). *Beyond the Blueprint.* Washington, DC: National Academy Press.

Rossi, P., Freeman, H., & Lipsey, G. (1999). *Evaluation.* (6th ed.). Thousand Oaks, CA: Sage Publications.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P., with Barnett, W. S., & Epstein, A. S. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27* (Monographs of the High/Scope Educational Research Foundation, 10). Ypsilanti, MI: High/Scope Press.

U.S. Department of Education. (1995). *National Goals 2000 Panel Report.* Washington, DC: Author.

U.S. Department of Health and Human Services (USDHHS). (1990). *Head Start research and evaluation: A blueprint for the future.* Washington, DC: Author.

U.S. Department of Health and Human Services (USDHHS). (1993). *Creating a 21st century Head Start. Final report of the Advisory Committee on Head Start Quality and Expansion.* Washington, DC: Author.

U.S. Department of Health and Human Services (USDHHS). (1995). *Charting our progress: Development of the Head Start program performance measures.* Washington, DC: USDHHS/Administration on Children, Youth and Families/Head Start Bureau.

U.S. Department of Health and Human Services (USDHHS). (1998). *Head Start program performance measures. Second progress report.* Washington, DC: USDHHS/Administration for Children and Families/Administration on Children, Youth and Families/Research Demonstration and Evaluation Branch and Head Start Bureau.

U.S. Federal Register. (1996, November 5). *Program Performance Standards for the operation of Head Start programs by grantee and delegate agencies* (pp. 57210–57227). Washington, DC: U.S. Government Printing Office

U.S. General Accounting Office (GAO). (1997, April). *Head Start: Research provides little information on impact of current programs* (Report No. GAO/HEHS-97-59). Washington, DC: Author.

U.S. General Accounting Office (GAO). (1998, June). *Head Start: Challenges in monitoring program quality and demonstrating results* (Report No. GAO/HEHS-98-186). Washington, DC: Author.

U.S. Office of Management and Budget (OMB). (1994). *The Government Performance and Results Act.* Washington, DC: Author.

Westinghouse Learning Corporation. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (Vols. 1–2, Westinghouse Learning Corporation and Ohio University report). Washington, DC: Clearinghouse for Federal, Scientific, & Technical Information.

Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *The Future of Children, 5*(3), 51–75.

Zill, N., Resnick, G., & McKey, R. H. (1999). *What children know and can do at the end of Head Start and what it tells us about the program's performance.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Albuquerque, NM, April.

The Long-Term Benefits of Head Start Study: Valuable Guidance

Mary Lou de Leon Siantz
Professor
School of Nursing
University of Washington

Head Start was created to strengthen the abilities of disadvantaged children and families and thus break the cycle of poverty. From its early years through the present, Head Start has existed to serve the most impoverished among the poor. It was founded on the premise that all children share certain needs and that the children of low-income families can benefit from, and should have access to, a comprehensive developmental program to meet those needs. Since its origins during the War on Poverty in the 1960s, Head Start has flourished and has impacted the lives of millions of low-income children and families.

Head Start embraces two major goals. The first is to provide disadvantaged children with the skills they need to be successful not only in school but also in life. The second goal is to assist families who participate in Head Start to become self-sufficient. To meet these goals, Head Start has had four components that focus on education, health, parent involvement, and social services. This approach has made Head Start not just a unique educational program but also a comprehensive child and family development program. However, does Head Start work? If so, how can Head Start continue to meet the needs of a changing population in the twenty-first century? What research is needed to strengthen the Head Start program to meet the needs of an increasingly multicultural population?

The findings of the Long-Term Benefits of Head Start Study provide evidence of long-term benefits extending into adulthood and suggest the importance of using a curriculum based on child development theory. The study findings, however, point up (1) the need for conducting more comprehensive assessments of family background in long-term studies of Head Start effects and (2) the importance of using random assignment to study groups and effective data analysis models in studies employing a quasi-experimental design to assess specific program effects.

CULTURALLY COMPETENT RESEARCH

In the twenty-first century, an increasingly multicultural preschool population will challenge Head Start researchers to become more culturally competent in their research. Head Start programs will be serving populations representing widely diverse ethnocultural and ethnolinguistic backgrounds. It will be important to examine the factors that contribute to the healthy development and learning of children and families from varied cultural and linguistic backgrounds.

Researchers will need to consider historical circumstances, such as the waves of immigration that have brought various subgroups of Head Start children and families, and data analyses will need to examine within-group variation. This is particularly true of studies involving Head Start participants who are Hispanic Americans. Within this broad group, Hispanic subgroups will need to be clearly identified (e.g., Mexican, Puerto Rican, San Salvadorean, Guatemalan). It will also be important to determine if the children are first-, second-, or third-generation Americans, because knowing their level of acculturation will help us to better understand what types of programs they and their families need and what types of long-term effects to expect. We will need to give more attention to bilingualism and biculturalism as they manifest themselves in families and in the broader society; we will also need to consider racial or ethnic discrimination and intergroup relations, especially as they affect the values, adult role models, and peer groups of Head Start children.

Research on ethnically diverse Head Start children must investigate factors that contribute to differences in outcomes within groups, not just between groups. For example, variables that explain school-achievement differences *within* certain minority groups are sometimes different from those that explain school-achievement differences *between* minority groups.

Research that builds on a multicultural framework and strengthens the curriculum with respect to multiculturalism can help us to address the unique contexts of the increasingly diverse preschool population in Head Start programs and the need for culturally competent research. Such research will be important in working toward the U.S. Department of Education's Goals 2000 objective of developing children's cognitive, social, and communication competencies so they can function effectively in a multicultural society and economy.

MODELS FOR FUTURE RESEARCH

Conceptual models that frame future research should examine not only risk factors but also positive outcomes in terms of the protective or compensatory conditions that enable children to thrive in spite of difficult life circumstances. Ethnographic studies should be embedded in longitudinal research of Head Start children. Despite having small samples that may not statistically represent the various populations, such studies can provide rich interpretations of the processes that more traditional quantitative research methods can highlight but not probe in depth. In particular, using focus groups of parents, staff, and Head Start teachers would allow the collection of contextual information in an informal, spontaneous format.

HEAD START'S COMPONENTS

Head Start has always been a comprehensive program that considers not only the individual child but also the child's ecological environment. This has meant meeting each child's health, dental, nutritional, and developmental needs and recognizing that the child develops within the family context. One unique characteristic of Head Start's model has been the commitment to providing a program of age-appropriate child development activities along with services that enable children's families to become self-sufficient. Thus, providing parents with education, social services, opportunities for economic advancement, and other family supports and seeking parents' maximum participation in their child's program have been an integral part of Head Start. This has empowered parents to be systematically and meaningfully involved in decisions affecting their child, their lives, and their programs. Helping families to establish linkages to the local community has been thought to be crucial to the success of children and families.

These interrelated components have always been important to Head Start's success. We must continue to investigate how these components can best predict long-term outcomes for the child and family, and we must also identify ways to establish research partnerships with poor, multicultural parents. Such partnerships will prepare parents to work with their children and teachers, will strengthen relationships between parents and local schools and communities, and will increase parents' support and commitment to participating in research when appropriate.

TEACHER AND STAFF PREPARATION AND DEVELOPMENT

Providing for teacher training and staff development is an important part of assuring delivery of quality Head Start programs at the local level. Particularly for communities with limited resources for training, we need to develop and test distance-learning techniques that include internet-based access to learning resources, teleconferencing, and self-directed CD-ROM or audiocassette training courses. This training for Head Start teachers and staff should cover a wide range of topics, including integration of research findings with teaching methods.

Research is needed to determine if strengthening teacher and teacher-assistant training as well as center-staff development will improve the long-term benefits of Head Start for preschoolers and their families. Also evaluation research is needed to assess the impact of teachers' academic preparation and experience on long-term benefits for children. New and innovative methods will need to be developed and tested to integrate Head Start research findings into the basic teacher-preparation through the Child Development Associate (CDA) program and continuing education. Research is also needed to investigate practical ways that staff development and training might be better integrated with day-to-day program activities.

CONCLUSION

The Long-Term Benefits of Head Start Study provides valuable guidance for establishing priorities for future studies. It should help us to develop strategies for research aimed at enhancing the effectiveness of the current program design by identifying its long-term benefits. The LTBHS study also furnishes evidence of the challenges that await researchers who work with poor and disenfranchised families. Such research is labor-intensive, statistically challenging, and costly, if scientific integrity is to be maintained. The study further suggests that researchers must be not only methodologically prepared but also culturally competent. The study underlines the fact that research is not a priority for poor families. Thus, particularly in longitudinal studies, families will best form partnerships with researchers who are sensitive to the culture, family, and community constraints.

Head Start is about to enter a period in which it will continue to serve as a national laboratory for best practices in early childhood and family support services within low-income, multicultural communities. A new generation of funded, culturally competent research is needed to identify innovative ways to produce long-term positive outcomes in the next millennium. Forming partnerships with multicultural families and communities will be crucial to the success of such research. An investment in America's children, who are of increasing ethnic diversity, is needed now more than ever to help the children of poverty develop the social and cognitive competence to become the law-abiding citizens, productive workers, nurturing parents, and leaders of the twenty-first century.

The Futures of Head Start's Children

Sheldon H. White
John Lindsley Professor of Psychology in Memory of William James
Department of Psychology
Harvard University

In 1994, I had the opportunity to assist in the completion of a an important study of Head Start's long-term effects on children's lives. With characteristic perspicacity, in this study, the High/Scope Educational Research Foundation had confronted the question of the hour about Head Start. It had found a significant number of 22-year-old graduates of Head Start programs and looked at how well they were doing in society. These young adults had been children in programs High/Scope had worked with in the 1970s under the auspices of the Head Start Planned Variation program. Now High/Scope had matched them with other, similar young adults who might have been, but were not, admitted to High/Scope's Head Start programs as children. High/Scope had collected information about the life circumstances of the two groups, and it was preparing to examine how the Head Start graduates, as compared with the non–Head Start adults, were faring in their lives.

It was likely that the analysis of the data would present problems. The Head Start sample had been established in the 1970s, but the comparison group had been put together years later by retrospective matching. It is very difficult to match two groups of subjects in this way, and one could expect that the matching would be imperfect. There are statistical procedures to correct for imperfect matching, but the issues that arise in using such procedures are complex and argumentative. Because of the potential importance of this study, a private foundation supported the appointment of an advisory panel on methodology and statistics, composed of Mark Appelbaum, Robert Boruch, Larry V. Hedges, and myself. There was a little division of labor among us. Appelbaum, Boruch, and Hedges are expert statisticians. I am not, but I have a long history of work with Head Start evaluations, including, from time to time, interpretations of what the evidence from such studies might mean in policy terms (e.g., White, 1971; White, Day, Freeman, Hantman, & Messenger, 1973).

In 1994 and 1995, the advisory panel made several trips to Ypsilanti to meet with Sherri Oden, director of the Long-Term Benefits of Head Start (LTBHS) study; David Weikart, president of High/Scope; Lawrence Schweinhart, chair of High/Scope's research division; and Mei-yu Yu and Benjamin Ben-Baruch, data analysts at High/Scope. We reviewed the data that had been collected in the LTBHS study, including emerging evidence indicating that the two comparison groups of the study were indeed not perfectly matched. Two technical advisors who could help with statistical analyses of the data were found: Sue Marcus of the University of Pennsylvania, and Yu Xie of the University of Michigan. Ultimately, Marcus and Xie would conduct two complete and independent analyses of the study's data, each correcting in a different way for the imperfect matching of the two groups.

The question of the hour, for some people, was whether genuine Head Start programs significantly change children's lives, bringing about such benefits as more success in school, less dependency, less delinquency, a better employment history, fewer unplanned pregnancies, or less drug use. Obviously, this is an interesting question, though not everyone agrees that it is an appropriate or reasonable one to ask when seeking to make an overall judgment about Head Start. Is it meaningful to ask of any preschool program that it produce measurable benefits in adolescence and adulthood? Why should anyone expect that a single good year in a 4-year-old's life, however wonderful, should make a large difference in the child's future fate?

The "question of the hour" has grown upon us by a process of political escalation. It is important to understand that Head Start has always cast a large and somewhat melodramatic shadow in the political arena. Visit any Head Start center and you will find a pleasant, low-key environment in which small children play in the midst of pleasant and helpful adults—a preschool in which the children are poor and often not English-speaking. Head Start is a good deal more complicated than an ordinary preschool, but it looks simple and straightforward and ordinary. Look upward, toward the state and federal political arenas in which authorizations, resource allocations, and program guidelines are created, and you see ideology, drama, and broad philosophical questions being debated on the basis of small and not-so-secure studies.

In the past 30 years, again and again, individuals on the political left and right have brandished the latest Head Start research to try to deliver a quick, punchy political thunderbolt to say once and for all that Head Start "works" or "doesn't work." These totalitarian claims have done much to sow confusion in public discussions of Head Start. Over the 30-plus years of the program's life, the meaning of the verb "works" has gradually escalated.

We have known since the beginning that children who participate in Head Start programs show clear and immediate gains on tests of cognitive development or school achievement. The evidence showing this is massive and comes from widespread sources. Some argue that these findings present a complete and sufficient justification for Head Start.

But these early benefits recede as the children grow older. During the third- and fourth-grade years, the test scores of groups of Head Start children and comparison children move together. Some say that Head Start's benefits "fade out" by the third or fourth grade—though, in fact, no one is completely sure what the convergence of the test scores means:

- *Perhaps Head Start's benefits are truly evanescent.* Possibly, the programs give children bits of knowledge and skill that elevate their performance briefly, but in time these marginal bits of knowledge and skill are used up.
- *Perhaps Head Start's benefits are squelched.* Schoolteachers may, intentionally or unintentionally, act against the changes in children that are produced by Head Start. Cognitive development, as much as theorists sing hosannas to it, does not always make children easier to deal with; sometimes it makes them more intellectually demanding, critical, "show-offy," and "fresh." In poor neighborhoods, where resources are in short supply, overworked teachers may put down the qualities given to children by Head Start.
- *Perhaps Head Start's effects are deflected by new life-course challenges.* Middle-class schools present a strange environment to the poor or minority child, and the child has about 4 years to find his feet in that environment before the "fourth-grade slump" sets in (Chall, 1983). As children reach the upper-elementary grades, schools turn from decoding and number facts toward serious reading and mathematics, and some benefits of Head Start might get lost as children confront these fresh demands. Similarly, as children progress into the school years, new demands on their social competence appear. Just outside the schoolroom, class and ethnic confrontations appear as a small-fry politics of childhood begins to open up in the fourth and fifth grades (Phinney & Rotheram, 1987).
- *Perhaps the appearance of fade-out is illusory.* The tests customarily used in Head Start studies are psychometrically sound, but they have never rested on much theory or depth of understanding. They are quick, pragmatic, rough-and-ready group-selection instruments. Perhaps Head Start produces changes in its children that conventional psychometric instruments are unable to detect. Possibly there are psychometrically hard-to-detect benefits of Head Start that manifest themselves as children try to meet the life-course challenges of the later years.

Evidence to date says there is something to the fourth hypothesis, that some of what looked like fade-out in the earlier studies must be illusory. From a number of sources, we have evidence that long-term benefits to graduates of early childhood programs can be detected in the years after "fade-out." But the programs for which such long-term benefits have been found have all been model programs—experimental or demonstration programs, often situated in or near universities, comparatively well staffed and well funded—not Head Start programs. Some have nevertheless argued that the findings of such long-term studies prove that Head Start "works." Others have argued that the data of the model programs prove nothing about Head Start. The LTBHS study being discussed in this report addresses the question of the hour—whether findings like those obtained from model programs can be obtained from genuine Head Start programs. This is an inevitable next-step question for researchers—although, again, its meaning for judgments about Head Start as a program are somewhat arguable.

Chapter 7 of this report gives a careful, itemized review of the LTBHS study's findings. Using a noisy and relatively inefficient research design, conservatively interpreted, the study found evidence that positive long-term benefits of children's participation in Head Start can be detected when the children have grown up and are 22 years of age. The difficulties of the design, it should be noted, have nothing to do with the quality of the researchers' work—which, in my view, has been admirable. Given the practical, political, and financial problems of conducting studies of human development in everyday-life settings, it is impossible to achieve the scientific standards set for treatment-comparison studies conducted in laboratory environments.

Some friends of Head Start will find the LTBHS findings satisfying, as a modicum of evidence testifying that Head Start "works" even according to the standards of the most demanding contemporary political debates. But other friends of Head Start will find these findings frustrating, because they seem to extend the Alice-in-Wonderland escalation of such political discussions. Looking at the matter either from a policy perspective or from the standpoint of scientific feasibility, I do not believe that Head Start's meaningfulness as a program can be given an up or down judgment according to whether it demonstrably establishes success and happiness in adult life.

The contemporary search for long-term effects of genuine Head Start programs will go on. Such inquiries are, of course, logical and inevitable, given what we now know to be true for model programs. But I hope that political discussion of Head Start as a program will now move beyond the 1960s mindset in which people imagined that our relatively narrow and weak research technologies could somehow be deployed to arrive at sweeping, summary evaluations about Head Start as a national program. We will need to construct a picture of what Head Start is and does through a mosaic of small studies, as an archaeologist constructs an image of an ancient village out of fragmentary observations and inferences. Chapter 7 of this report offers recommendations for future Head Start research that seem like excellent rubrics for a program of such studies.

References

Chall, J. S. (1983). *Stages of reading development* New York: McGraw-Hill.

Phinney, J. S., & Rotheram, M. J. (Eds.). (1987). *Children's ethnic socialization: Pluralism and development.* New York: Sage.

White, S. H. (1971). The National Impact Study of Head Start. In J. Hellmuth (Ed.), *The disadvantaged child* (Vol. 3). New York: Brunner/Mazel.

White, S. H., Day, M. C., Freeman, P. A., Hantman, S. A., & Messenger, K. P. (1973). *Federal programs for young children: Review and recommendations* (Vols. 1–3). Washington, DC: U.S. Government Printing Office.

The LTBHS Study's Contributions to the Future of Head Start

Edward Zigler
Sterling Professor of Psychology
Director, Bush Center in Child Development and Social Policy
Yale University

Sally J. Styfco
Associate Director, Head Start Research Unit, Bush Center in Child Development and Social Policy
Yale University

Head Start is a complex program, and so is its evaluation. The LTBHS study perfectly illustrates this point. The investigators are among the best and most experienced scientists in the field, they were advised by a panel of leading scholars, and they tapped the wisdom of statistical experts to create fitting analytic methods and alternatives. The arduousness of their task is witnessed by the literature's paucity of sound longitudinal research on Head Start. In a review of all post–Head Start evaluations, the General Accounting Office examined over 600 studies and found only 22 that met their criteria of valid and timely approaches to the issue of program benefits. The Oden, Schweinhart, and Weikart study would certainly have been on the short list of acceptable investigations. The study described in this book represents the best work conducted to date on Head Start's impact on the life course of people it once served.

For many years evaluations focused on the basic question *Does Head Start work?* In its blueprint for future research, the Advisory Panel for the Head Start Evaluation Design Project urged that the field advance to asking *What services work best and for whom?* The LTBHS study is a step toward this next generation of research. Its design, implementation, and results expose just how complicated this type of assessment can be. It forewarns that there will be no easy answers to the panel's rallying question.

The LTBHS findings show that the effects of Head Start vary among programs and among subgroups of participants. Site effects were pronounced, even though only one component of Head Start's multifaceted program was examined. Preschool education services were found to vary in delivery and impact, suggesting that such differences will likely characterize Head Start's other core components. Differential effects were also discovered for subgroups of children. Mediating factors included gender, racial/ethnic group membership, and family education and socioeconomic status. While these results begin to address the question of which services work best and for whom, they indicate that this is a multiple-choice question. The search for answers must be conducted both within and outside of Head Start.

Inside the program, future investigators must include process or qualitative evaluations of the classrooms they engage. Outside features must include detailed demographics of the children and families who come to Head Start. And, as this study's authors recommend, we must look not only at Head Start but also at the resources and opportunities that are available to the programs and participants in their respective communities.

The results of the LTBHS study are not as robust as the investigators had hoped they would be. Nonetheless, they are extremely valuable in that they confirm many of the suggestions, indications, trends, and tendencies apparent throughout the sizable body of Head Start literature. The pattern of earlier findings is repeated in the LTBHS study and, given the study's methodological strengths, this substantiates our hunches enough that we can now speak with greater confidence.

One hunch that is supported is that early intervention programs appear to benefit girls more than boys. This finding has surfaced in many (but not all) studies of Head Start and of other early childhood or family-support programs. It was evident, even into adulthood, among the graduates of the High/Scope Perry Preschool program. Whether girls are more resilient or boys are more vulnerable to the effects of poverty—or whether the two genders are differentially open to helping services—are important questions. To reach answers that will ultimately guide the development of more effective programs, future studies should at least be sensitive to these differences.

Another idea given weight by the LTBHS findings is that existing studies, this one included, generally underestimate Head Start's effects. Part of the reason is the nature of the quasi-experimental designs that characterize much of the research. The rest is the fact that Head Start really does serve the poorest of the poor. Program guidelines urge local centers to enroll children and families who are most in need of services and therefore stand to benefit the most. (This is why the control groups were initially of higher socioeconomic status than the program groups.) The Head Start participants thus begin the farthest away from the intervention's goals and have the most ground to cover to achieve them. When participant progress is assessed against middle-class standards, the program appears minimally effective; when their progress is assessed against where they were when they started, the program's benefits are meaningful.

That there is value in continuing services after the Head Start experience is another view corroborated by the LTBHS study. The largest program benefits were reaped by children who attended the model Head Start sites in Florida. Most of these children went on to elementary schools that had Follow Through programs, where the same model curriculum was extended through third grade. Reports of other programs, including the Abecedarian Project and the more Head Start–like Chicago Child-Parent Centers, also indicate that longer services lead to better outcomes. The original High/Scope Perry Preschool program was attended for 2 years by most of the students in the study, and the curriculum was brought into their homes through teacher visits and parental involvement. All of this proves to us that the principle of developmental continuity is important. It is fruitless to search for quick-fix solutions to languishing problems of poverty. We must stop advocating for Head Start as a stand-alone program and begin adapting its proven approach into the grade schools that at-risk children attend.

Finally, if there was ever any doubt that quality matters in children's services, the LTBHS findings constitute proof. Children who attended the higher quality Head Start classrooms had better outcomes, including higher GPAs and fewer criminal convictions. The curriculum delivered in these classes was well planned and well implemented, and the teachers were relatively better trained. While these findings may point to nothing new, they do point to the direction quality-improvement efforts in Head Start must take. They also corroborate the need for the enhancements that are now under way, thanks to the rapid response of the Clinton administration to the recommendations of the Advisory Committee on Head Start Quality and Expansion.

The LTBHS study makes a major contribution beyond confirming our theories and educated guesses. It is the closest we have to a replication of the longitudinal study of the High/Scope Perry Preschool program. This study is widely cited as evidence that high-quality early childhood programs are cost-effective and can have lasting benefits. Hopes that these programs can improve educational and career outcomes and reduce welfare usage and crime sprang from this classic study. Yet the findings were isolated to a rather small group who attended a single program many years ago. One could only surmise that similar results would occur outside of Ypsilanti, Michigan. The LTBHS study suggests that they may. The treatments were not exactly the same, but the LTBHS study's Model Head Start group in Florida did receive the High/Scope curriculum. As adults, features in their academic and criminal histories resembled those of the Perry preschoolers. Although cost/benefit analyses were not conducted, the LTBHS findings confirm at least some of the original Perry results.

Some errors of the past are repeated in the design of the LTBHS study, so it does not provide the groundwork to help us correct them, as it does for some other methodological problems. The study is limited to assessing the impact of preschool education services and does not address the benefits of Head Start's other components—the health, social services, or parent-involvement features. Of course, the design of the evaluation was limited by the scope of the original data set, which was collected before scientists showed much interest in such things. How well these other components were implemented is unknown. The study participants attended Head Start before there were program performance standards, so the quality of these services may not have been up to par. This thought cannot be proved or disproved, given the study's retrospective design—a problem that also plagued past research. Oden and colleagues did the best they could with what they had, but this weakness underscores the need for *prospective* longitudinal studies. Such work requires funding commitments and investigators who have the diligence and stamina modeled by the LTBHS team.

A prize from the LTBHS study is the encouragement it provides. The investigators show us that the obstacles to longitudinal research can be overcome and enough good data collected to allow reasonable conclusions. The findings confirm that we are on the right track in deploying comprehensive interventions and advocating for high-quality services. They impart a needed dose of reality to our sometimes too-high hopes. Head Start does a good job in preparing children for school, but it cannot change their whole lives. By adulthood, some are better-off because they went to Head Start, some are not. By conducting longitudinal studies such as this one, we raise expectations that Head Start should have lasting benefits. (Otherwise, why would we be looking?) But by not conducting them, we may never come to grips with what at-risk children need to succeed after they leave Head Start. The LTBHS study inspires this search. The investigators are to be congratulated for an outstanding piece of work.

Into Adulthood

A Study of the Effects of Head Start

Sherri Oden, Lawrence J. Schweinhart, David P. Weikart
with Sue M. Marcus and Yu Xie

Into Adulthood describes the background, design, and findings of the Long-Term Benefits of Head Start (LTBHS) study. The study followed up children at two sites, one in Florida and one in Colorado, 17 years after their participation in the 1970–71 Head Start program. Prepublication commentators had this to say about this report and the LTBHS research effort:

We will need to construct a picture of what Head Start is and does through a mosaic of small studies, as an archaeologist constructs an image of an ancient village out of fragmentary observations and inferences. . . . this report offers recommendations for future Head Start research that seem like excellent rubrics for a program of such studies.

SHELDON H. WHITE
HARVARD UNIVERSITY

As we educate the public about what matters in early childhood education, parents and policymakers will demand that our research help them understand the relationship between program quality and later child outcomes. Toward that end, we will need more studies like the LTBHS study—ones that are longitudinal in nature, extend beyond the elementary school years, and incorporate sufficient information about the nature of the programs studied.

SUZANNE M. RANDOLPH
UNIVERSITY OF MARYLAND—COLLEGE PARK

The LTBHS study . . . furnishes evidence of the challenges that await researchers who work with poor and disenfranchised families. Such research is labor-intensive, statistically challenging, and costly, if scientific integrity is to be maintained.

MARY LOU DE LEON SIANTZ
UNIVERSITY OF WASHINGTON

A prize from the LTBHS study is the encouragement it provides. The investigators show us that the obstacles to longitudinal research can be overcome and enough good data collected to allow reasonable conclusions. The findings confirm that we are on the right track in deploying comprehensive interventions and advocating for high-quality services.

EDWARD ZIGLER, SALLY J. STYFCO
YALE UNIVERSITY

BEST COPY AVAILABLE

HIGH/SCOPE
PRESS
[illegible] Research Foundation
[illegible]
[illegible]

ISBN 1-57379-089-3
90000>
9 781573 790895